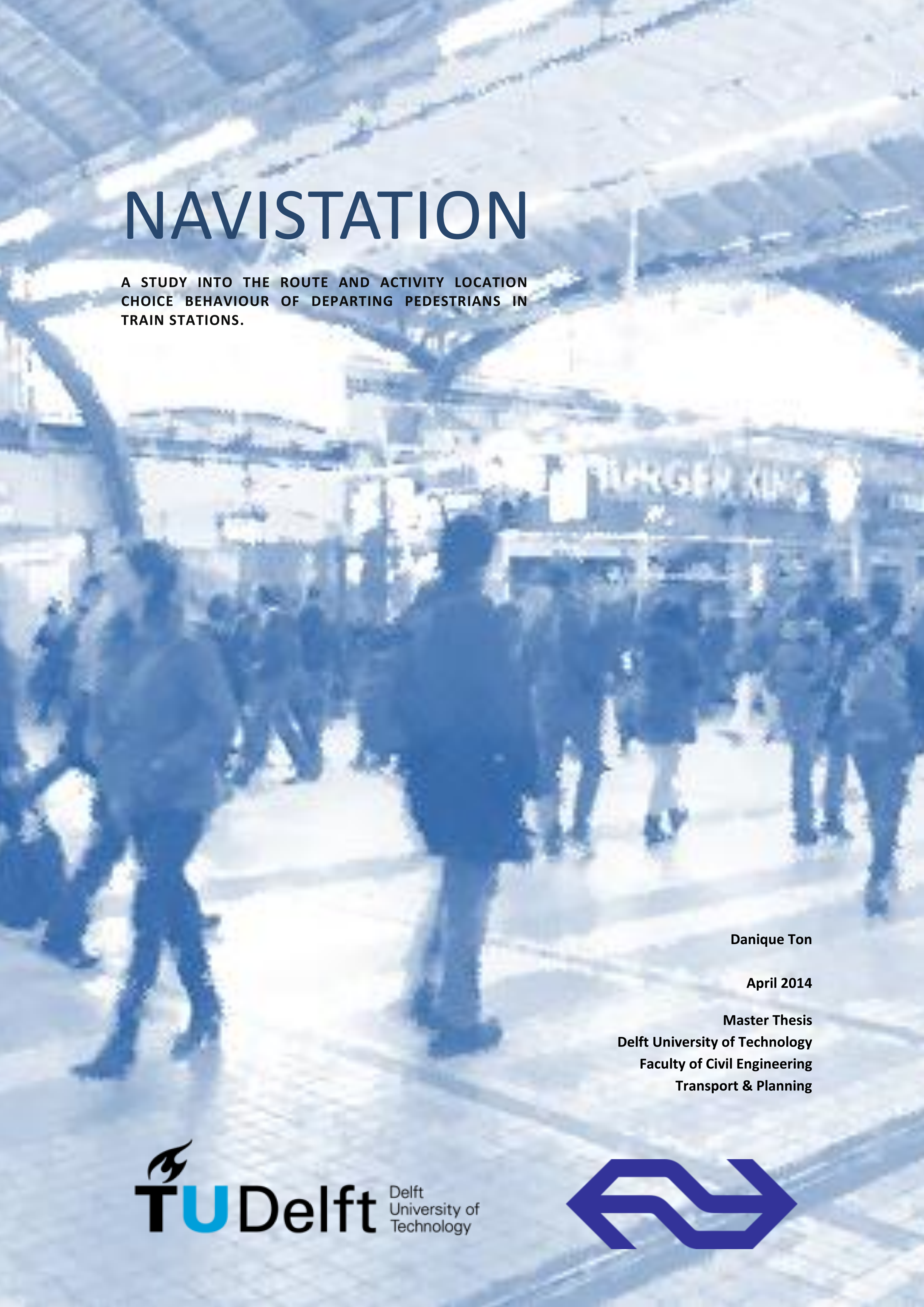

NAVISTATION

A STUDY INTO THE ROUTE AND ACTIVITY LOCATION CHOICE BEHAVIOUR OF DEPARTING PEDESTRIANS IN TRAIN STATIONS.

Danique Ton

April 2014

Master Thesis
Delft University of Technology
Faculty of Civil Engineering
Transport & Planning

NAVISTATION

A STUDY INTO THE ROUTE AND ACTIVITY LOCATION CHOICE BEHAVIOUR OF DEPARTING PEDESTRIANS IN TRAIN STATIONS.

Danique Ton
1511076

Graduation committee:

Prof.dr.ir. S.P. Hoogendoorn – *Chair*

Delft University of Technology
Faculty of Civil Engineering – Transport & Planning
S.P.Hoogendoorn@tudelft.nl

Ir.drs. J.P.A. van den Heuvel MBA - *Daily Supervisor – NS Stations*

NS Stations
Vastgoed & Ontwikkeling
Jeroen.vandenheuvel@nsstations.nl

Dr.ir. W. Daamen - *Daily Supervisor – TU Delft*

Delft University of Technology
Faculty of Civil Engineering – Transport & Planning
W.Daamen@tudelft.nl

Dr. E.J.E. Molin - *External Supervisor*

Delft University of Technology
Faculty of Technology, Policy and Management – Transport & Logistics
E.J.E.Molin@tudelft.nl

Ir. P.B.L. Wiggenraad - *Graduation Coordinator*

Delft University of Technology
Faculty of Civil Engineering – Transport & Planning
P.B.L.Wiggenraad@tudelft.nl

Contact author: danique.ton@nsstations.nl or danique29@hotmail.com

Cover photo made by the author

A request for appendices can be sent to the author

PREFACE

This thesis report shows the results of my graduation project on the activity location and route choice behaviour of departing pedestrians in train stations. This thesis is fulfilled as part of the Master Civil Engineering at Delft University of Technology and produced in cooperation with NS Stations.

I would like to express my gratitude to all the people who have helped me create this thesis. First of all, I would like to thank my supervisor at NS Stations, Jeroen van Heuvel, for his constant and deeply appreciated advice and help. I enjoyed all the conversations we had about research, work and life. Thank you for introducing me to your passion for trains, it made me realise I am the same! Also, I would like to thank my supervisor at the TU Delft, Winnie Daamen, for her detailed feedback and support throughout the entire project. Furthermore, I would like to thank the other members of my graduation committee – Serge Hoogendoorn, Eric Molin and Paul Wiggenraad – for their encouragement, support and feedback.

All colleagues at NS Stations, thank you for the time and effort you invested to get to know me. The conversations about work, the future and life helped me deciding about my own future. Thank you for making me feel part of the team. It was (and will be) wonderful working with you!

Furthermore, I would like to thank Kim Hermansen from Blip Systems for the fruitful discussions about SMART station and the data; John Tuithof from NS Reizigers for providing me with the data on the train processes at Utrecht Central station; Noor Scheltema for her help with the pictures of Utrecht Central station; Rachel Halfpenny for her feedback on the final report; my fellow students at the university for their food for thought during coffee breaks; my friends for the fun times and good discussions and my family for their support and trust in my study and thesis. At last, I would like to thank Rudolf for his understanding, trust and patience during my entire studies and this research.

Danique Ton

Delft, April 2014

EXECUTIVE SUMMARY

The number of train passengers increased over the past years. This result is that train stations get more crowded. In addition, the use of the stations changed into a place to stay and meet people. These changes ask for a good management strategy. However, in order to develop a good strategy it is necessary to understand how pedestrians behave and make their choices. Therefore, the objective of this research is to understand and predict how pedestrians choose their route and activity locations in a station area from entering the station building to boarding the train. This goal is reflected in the main research question:

'Which factors and processes influence the route and activity location choice behaviour of pedestrians in train stations from entering the station building to boarding the train and to which extent?'

CHOICE BEHAVIOUR OF DEPARTING PASSENGERS IN TRAIN STATIONS

Departing pedestrians make several choices in the train station. The most fundamental choices are the activities to perform and the route to walk. Related to the choice for activities are the hierarchy amongst the activities, the sequence of the activities and the location on which to perform the activities.

The factors identified were divided over four main groups: personal, system, public transport and external. The personal factors were divided into general personal, trip factors and learning processes. The system factors were divided into location factors and route attributes. One factor, the familiarity of a person, was identified by many studies as a significant influence on the choice behaviour of pedestrains. This means that a familiar pedestrian makes choices in a different way than an unfamiliar person. Several frameworks were found in literature that identified relations between the choices made by the pedestrian. These frameworks identified two levels that in pedestrian choice behaviour: before trip and during trip. A pedestrian makes plans before the trip (usually seen as the planning of activities) and executes a certain plan (including adaptions) during the trip. Here he determines the location, route and schedule of the activities. These findings were used in this research in the creation of the theoretical framework for this study.

Due to the significant difference expected in the choice behaviour of familiar and unfamiliar pedestrians, two frameworks were introduced in this study (familiar versus unfamiliar passengers). These frameworks include decisions made on two levels: before trip and during trip. The familiar passenger knows the station and can therefore plan further ahead than the unfamiliar passenger. Therefore the familiar passenger plans all aspects of his trip on beforehand. He then adapts the sequence, route and locations of activities when needed (could be due to e.g. pedestrian operations, later arrival time, impulse behaviour or a delayed train). The unfamiliar passenger is only able to plan the activities, assign importance to these activities and plan a basic activity sequnce before the trip. During the trip he will determine his actual activity sequence, route and locations.

Not all relationships identified in the framework, relating to route choice and activity location choice, can be quantified in this study. This means that a selection is necessary. This selection is based on two criteria. The first is the availability of knowledge. If the relationship is not researched or not quantified in research, it provides opportunities for this study. The second criterion is the ability of the SMART station data to capture the relationship.

SMART station is the method used to collect data. The concept of SMART station consists of tracking and counting pedestrians in the station building. The counting is carried out by infrared scanners at the entrances. The tracking is done by Wi-Fi and Bluetooth scanners. The scanners use the technologies available on e.g. smart phones to track pedestrians through the station building. Not all pedestrians carry a device that has either of the technologies enabled, therefore the infrared scanners are also used to calibrate the data collected.

The data collection method SMART station has some limitations regarding the data it collects. Due to privacy issues personal data are not registered. This means that relationships concerning socio-economic data cannot be captured by the data. In addition, the data are of a binary nature. This means that the method only registers if a pedestrian is present at the location or not. The result of this limitation is that the exact movements of the pedestrian cannot be registered. This results in the fact that not all the relationships, of which no knowledge is available, can be captured by means of SMART station data.

All relationships were assessed on the two criteria identified. This resulted in the following relationships that will be quantified in this study.

	Route choice	Activity location choice
Time spent in the station hall	X	X
Time of day or week	X	X
Distance	X	X
Travel time	X	X
Time table	X	X
Train operations	X	X
Visibility	X	
Orientation	X	X

CASE STUDY – UTRECHT CENTRAL STATION

The relationships are quantified by means of data collected at Utrecht Central station. The train processes are of major influence in the behaviour of departing pedestrians, as their final destination in the station is the train. Utrecht Central station consists of six platforms that can be divided into three typologies: final/start platforms, intercity platforms and mixed platforms. These platform types are different in the train processes present, therefore it can be expected that pedestrians behave different.

The data collected by SMART station are validated with several tests. It can be concluded that the dataset seems to be representative of the population of departing pedestrians at Utrecht Central station. The number of departures during the day and over the days is significantly equal. The distribution of time spent was as expected. The distribution of pedestrians over the platforms is a bit skewed. Direction Amsterdam Central station is most frequently taken, but this does not show in the data. This could be due to a lot of rescheduling of the train during the data collection period.

As mentioned before the data of SMART station are collected by Bluetooth and Wi-Fi scanners. It was expected that these technologies are used by different types of pedestrians. Bluetooth's most popular function is the wireless headset for making phone calls, whereas the main use of Wi-Fi is for accessing the internet. Therefore, several tests were carried out to see if these groups are really different. They are equally distributed over the day and during the day. However, the time spent distribution differs significantly for the two groups. Bluetooth users spent less time in the station building. Next to that, some differences arose in their presence in the shops. Especially regarding AH to Go and Tickets and Service. This means that the differences needed to be taken into account in the quantification process of the relationships.

CONCLUSIONS

The method used for quantifying the relationships is discrete choice modelling using the concept of utility maximisation. The underlying assumption is that the departing pedestrian chooses the alternative that provides the highest subjective utility. Two discrete choice models are estimated: route choice model and activity location choice model, that take the relationships selected into account.

Route choice consists of the entire route from entering the station building to boarding the train. For this model only the departing pedestrians heading straight to the platform are taken into account. If no activity is performed, the route choice can be estimated separately from activity location choice. The platform used for the estimation of the route choice model is 11/12.

Not all factors selected can directly be included in the model, these need to be translated to operational factors. Travel time and time spent are considered the same for this model. Time of day or week is reflected by peak and off-peak hours, which are a proxy for the familiarity of pedestrians. Familiar pedestrians travel more during peak hours and unfamiliar pedestrians during off-peak hours. The time table is translated to the stop location of the train; does the route provide direct access to the train or not. The train operations are reflected by the delay.

The results of the model estimation are visualised in Figure 1. The travel time (no differences were found for Wi-Fi and Bluetooth users), distance, orientation and stop location of the train provided significant influence on the route choice. A combined model could be estimated taking travel time, orientation and the stop location of the train into account. The travel time and distance are highly correlated and can therefore not be combined into one model. As the travel time is the most dominant factor, therefore this was preferred over distance. In the estimated model time spent is most influential on the choice behaviour of pedestrians. If a route provides direct access to the train on the platform, the probability for choosing that route increases. If this is not the case it decreases. In addition, the route located on the right side of the departing pedestrian has a higher probability of being chosen. Cars drive on the right, trains ride on the right and thus people walk on the right.

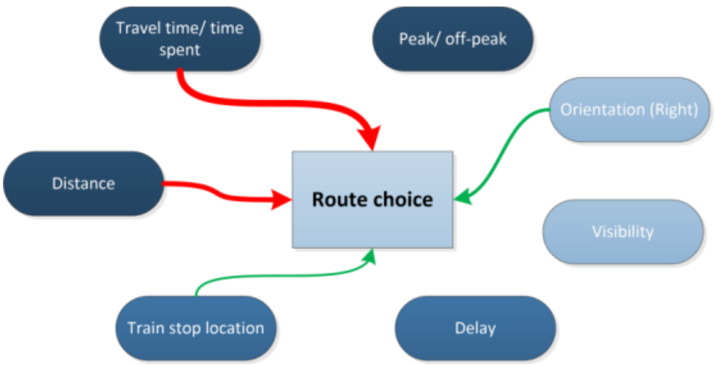

Figure 1: Factors influencing route choice

The peak and off-peak hours do not influence the route choice. This factor served as a proxy for the familiarity of the departing pedestrians. This means that regarding this proxy familiar and unfamiliar pedestrians choose their route the same way. The relationship between the proxy and familiarity however is not one on one, therefore it cannot be concluded that the familiarity of a pedestrian is of no influence on route choice. The fact whether a train is delayed does not influence the stair choice of the pedestrian. Next to that, the visibility of the stairs are of no influence on the stair choice.

Activity location choice is the choice for a certain location to perform an activity. In order to identify the factors and processes of influence on this choice, all departing pedestrians that perform that activity on either of the locations are taken into account. This means that people with multiple origins and destinations are included. The activity selected for the location choice model is buying a coffee. The locations included in the model are Starbucks Bruna and Starbucks BK.

Not all factors selected can be directly included in the activity location choice model, they need to be operationalised. Travel time is separated into two parts, travel time from entrance to location, and travel time from location to platform. The time of day or week is reflected as travelling during peak hours or during off-peak hours. The distance is reflected by two factors: the total distance and whether the pedestrian needs to make a detour to visit the location. The time table is reflected by two factors: the service type of the train and the platform of departure. Finally, the train operations are reflected by the delay of the train boarded.

The results of the model estimation are visualised in Figure 2. The travel time before (no difference were found in the model estimation for Wi-Fi and Bluetooth users), total distance, detour and orientation are of significant influence. Combining all significant factors into one model however is not possible. The distance is so dominant that the orientation and detour can be neglected when combined with this factor. The travel time is less dominant than the distance, and influences the location choice in the same way. The higher the travel time, the lower the probability for choosing that location. A detour in the route has the same effect. When the location is on the right side of the pedestrian regarding the entrance, the probability for choosing that location increases. However, the latter two factors are less important in the decision making than the total distance. This seems to indicate that the pedestrian optimises the entire route when choosing the location, especially aimed at the first part of the route.

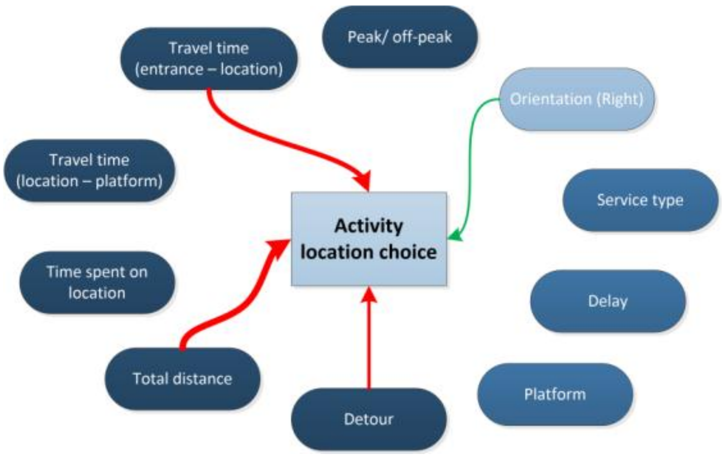

Figure 2: Factors influencing activity location choice

The peak and off-peak also turn out not to influence the location choice. Regarding this proxy the familiarity of the pedestrians does not matter for choosing a location. However, this conclusion cannot be derived as the relation is not one on one. The platform itself does not influence the location choice, the combination of entrance and platform (reflected by distance) does. The service type of the train that is boarded also does not influence the location choice. It does not seem to matter which type is boarded (intercity, sprinter or international train). The delay also does not influence the location choice.

RECOMMENDATIONS

This study could only quantify some of the relationships identified in the theoretical framework. Therefore, recommendations for future research include quantifying the factors and processes of influence on the choices before the trip, relations between choices before the trip, to what extent pedestrians plan their trip in advance and how pedestrians update/adapt their choices during the trip. Also a combined model could be estimated for route and activity location choice (super network). The concepts of regret minimisation and bounded rationality could be tested. Finally, the revealed data could be combined with stated preference data, to get a better understanding of the choice behaviour of pedestrians.

CONTENT

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The number of train passengers increased significantly over the past 40 years (Veenendaal, 2004). Now, every day over half a million people travel by train in the Netherlands (Bakker & Zwaneveld, 2009). Of these travellers, most visit the largest stations in the Netherlands: Amsterdam Central station, Utrecht Central station, Rotterdam Central station and Den Haag Central station. The growth in the number of train passengers in the Netherlands is accompanied by a change in the use of train stations. The train stations changed from a necessity for entering the train to a place to meet people and stay for longer time periods. In the stations more shops, waiting areas and extra services are present (Nio, 2012).

The change in the use of the station and the growing number of train passengers lead to the necessity of knowing and understanding how people behave in the station, especially in the largest train stations. When it is known how people behave, a strategy can be developed to manage the large amount of (crossing) pedestrian flows present in the station building. The basics of pedestrian behaviour can be captured by looking at the choices people need to make in the train station. Choices made in the train station are for example the locations at which to perform an activity and route taken (see Chapter 2). This comes down to deciding on how to navigate through the station area.

A collaboration between NS Stations (Netherlands Railways), NPC (RoyalHaskoningDHV) and BLIP systems led to the development of a measurement method that is able to track and count pedestrians in the station building: SMART station (see Chapter 3 for more information). It collects data on what many pedestrians do in the train station. The data collected with SMART station can be used to analyse the choices made by pedestrians in train stations. SMART station is installed at several stations in the Netherlands (e.g. Utrecht Central station).

The necessity of knowing how people behave in the station building can be partially fulfilled by knowing and understanding the choice behaviour of people in the station. This thesis therefore addresses the choice behaviour of pedestrians in train stations. The data was collected at Utrecht Central station. Utrecht Central station is the second largest train station in the Netherlands (after Amsterdam Central station) and the largest station equipped with SMART station.

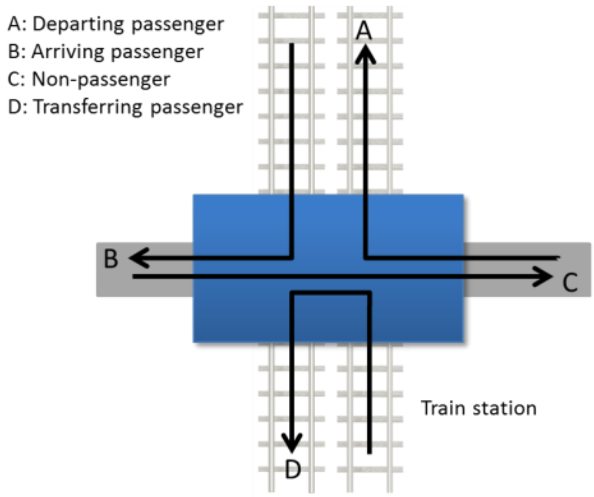

Figure 3: Passenger types in the train station

In the station building four types of passengers are present: departing passengers, arriving passengers, transferring passengers and non-passengers (see Figure 3). The largest flows present in stations the Netherlands are the arrivals and departures. Several studies addressed the behaviour of the arriving passengers in train stations (e.g. Lee et al. (2001) or Qi et al. (2008)). The research concerning departing passengers is limited. Some studies did address the station as a whole (e.g. Daamen (2004)), meaning that multiple passenger types are taken into account. The departing passengers do not arrive in large groups, but present a more constant flow. The departing passenger on the other hand provides a very interesting party related to activity choice behaviour. The departing passenger spends more time in the station building than the arriving passenger. This means that this passenger type has more time to perform activities and therefore provides an interesting group for research.

This thesis therefore aims to understand and explain how passengers make their choices in the station from entering the building to boarding the train.

1.1. RESEARCH OBJECTIVE AND RESEARCH QUESTIONS

The objective of this research can be summarised as follows:

The objective of this research is to understand and predict how pedestrians choose their route and activity location in station areas from entering the station building to boarding the train

The research is started by investigating the choice behaviour of departing pedestrians inside the train station and the factors and processes that influence them. These factors are identified through a literature review and observations from practice. Then the relationships between the choices made are identified. This leads to the creation of a theoretical framework on choice behaviour. This framework provides a qualitative insight in understanding how people make their choices. Via a selection procedure some of the relationships in the framework concerning route and activity location choice are identified for quantitative research. The other relations, which are already known in literature or which cannot be captured quantitatively by SMART station are left out. The selected relationships are researched by means of discrete choice analysis. This way the extent to which the relationships are present can be identified. This results in a better understanding and explanation of the route and activity choice behaviour of departing pedestrians in the station building.

In order to structure the objective, several research questions are defined. The main research question is:

‘Which factors and processes influence the route and activity location choice behaviour of pedestrians in train stations from entering the station building to boarding the train and to which extent?’

In order to answer the main research question several sub questions are identified. These are discussed below.

(1) Which factors and processes influence the route and activity location choice behaviour of departing pedestrians within train stations?
(2) Which relationships exist between route and activity location choice behaviour of departing pedestrians within train stations according to the state-of-the-art?
(3) To what extent is SMART station a suitable data collection method for collecting quantitative data on route and activity location choice behaviour of departing pedestrians in train stations?
(4) Which types of discrete choice models belonging to the logit family can be used to estimate models on route and activity location choice behaviour of departing pedestrians?
(5) Which factors have a significant influence on the route choice behaviour of departing pedestrians?
(6) Which factors have a significant influence on the activity location choice behaviour of departing pedestrians?

The first sub question is answered by presenting an overview of the factors and processes that influence activity location and route choice behaviour from the state-of-the-art and the state-of-the-practice. The question is specifically aimed at departing passengers in train stations, therefore the non-relevant factors and processes are left out. This question is addressed in Chapter 2. The second sub question addresses the relationships between the choices made in train station. This question is answered by presenting several usable frameworks identified in the state-of-the-art which cover these relations. This question is also discussed in Chapter 2.

It was decided beforehand that the method SMART station would be used to collect data on the choice behaviour of pedestrians. However, in order to know the trustworthiness and reliability of the method it needs to be determined whether it really is a suitable method. This is addressed in question three, which is answered in Chapter 3.

The method identified for analysing the data collected by the SMART station method is discrete choice modelling. It needs to be determined which types of models of the logit family can be used to capture route and activity location choice behaviour of departing pedestrians in station buildings. This is addressed in question four and is answered in Chapter 4.

The objective aims to understand and predict the factors and processes that influence route and activity location choice behaviour of departing passengers in train stations. Therefore two models are developed: a route choice model and a activity location choice model. Question five addresses the factors and processes of influence on the route choice model. This is discussed in Chapter 5. Question six addresses the factors and processes of influence on the activity location choice model. This is discussed in Chapter 6.

1.2. RESEARCH SCOPE

The scope of this research is determined in this section. The following aspects are addressed: the departing passenger, use of SMART station method, the location of the data collection (Utrecht Central station), the level of detail in choice processes and discrete choice analysis.

DEPARTING PASSENGERS

In the introduction it was stated that limited knowledge is available on departing passengers. Also the longer time spent by this passenger type in the station building is of relevance. More time in the station building means more time for performing activities. Therefore this group is of interest for research related to activity location and route choice behaviour. Figure 4 clarifies the differences between the arriving and departing passenger.

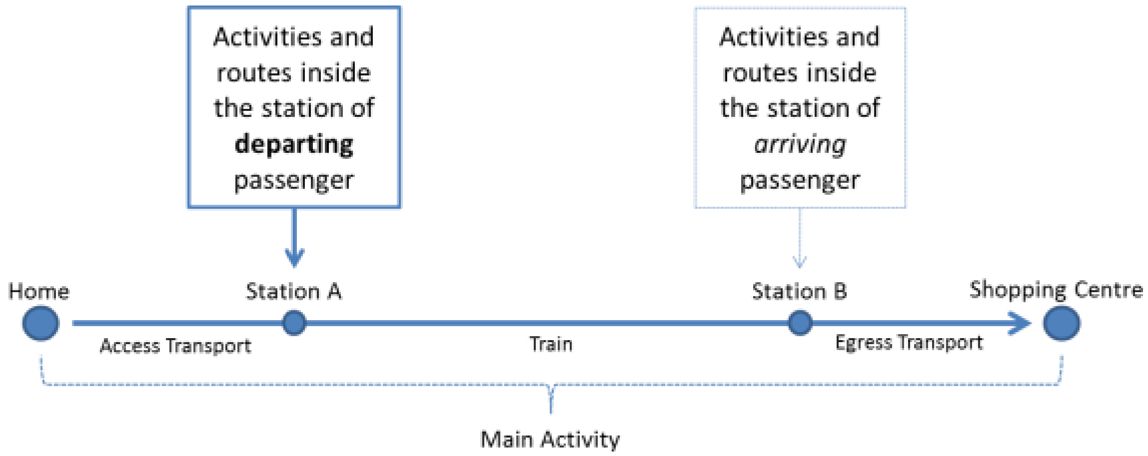

Figure 4: A trip made by a passenger

A passenger makes a trip to perform his main activity. In the example the main activity is to visit the shopping centre. This main activity is outside the scope of this research. The passenger first leaves his home to go to the station in order to access the train. He does this using access transport. In the station he needs to board his train. Also, he can perform activities and follow a route in the train station. This is where he is the departing passenger. He then boards the train and travels to the station at his destination. There he is the arriving passenger and can perform activities and routes inside the station. He then leaves by means of egress transport to the main activity.

As mentioned before, the departing passenger spends more time in the station compared to the arriving passenger. He therefore has more time to perform activities. A transferring passenger could also spend a long time in the station building, however the time tables at Utrecht Central station are such that the maximum transfer time is 15 minutes. The activities performed by the transferring passenger could thus be the result of compulsory time spent at the station. The departing passenger chooses himself when to arrive. In 60% of the cases the departing pedestrian uses a bike or walks to the station. The other 40% uses the bus or tram (NS, 2011). The frequency of these public transport services in Utrecht Central station is high. Therefore, this will not infer much compulsory time in the station building.

SMART STATION

It was decided beforehand that SMART station would be used to collect data on the behaviour of pedestrians. This means that for the quantitative research a selection must be made on relationships that can be tested by means of data collected by this method. It could be that other relationships provide more interesting information, but they cannot be tested by means of these data. See Chapter 3 for more information on SMART station.

UTRECHT CENTRAL STATION (LOCATION OF DATA COLLECTION)

Utrecht Central station was selected for collecting quantitative data using SMART station. This is the station with the largest coverage with SMART station scanners. Next to that, Utrecht is the second largest station in the Netherlands (after Amsterdam Central station). Also the range of possible activities to perform and number of locations in which this is possible provide added value. This makes Utrecht an interesting research location. When comparing the use of access transport to the station, Utrecht is comparable to, for example, Rotterdam Central station and Amsterdam Central station. Also the distribution of departing passengers over the day is comparable (NS, 2011). The number and range of activities is also large at these stations. This could mean that the research conducted at Utrecht Central station can also be used at other comparable stations. Note that Utrecht Central station is currently under construction.

In 1997 it was decided that the station area of Utrecht Central station should be expanded. Due to the number of passengers travelling via Utrecht Central station every day and the future prospects it was necessary to expand the station. The fastest way to do this, with respect to engineering works, would be to close down the entire station and rebuild it in one go. However, because Utrecht Central station is a major public transport hub in the Netherlands it would affect too much of the train traffic. It was decided that the station should be expanded in several phases. This means that the engineering works will last longer. The engineering works of the station building started in 2011 and are planned to be finished in 2016 (CU2030, 2013).

The data collection of this research was done from the 30th of August 2013 to the 19th of September 2013. A new building phase was finished on the 30th of September, therefore the lay-out of the station has already changed compared to the data collection period. The situation in the train station during this research period was relatively steady. This means that no changes were made to the configuration of the station and that the travellers were not interrupted by works in their activities. Next to that, this configuration was already present for a long period, meaning that travellers were already used to the situation. Therefore, the use of a station under construction in the data collection does not seem to have any negative implications. See Chapter 4 for more information on Utrecht Central station.

LEVEL OF DETAIL IN CHOICE PROCESSES

The use of SMART station for analysing the choice behaviour of departing pedestrians in train stations implies a constraint on the level of detail that is possible in the analysis. The SMART station method is not able to capture the exact steps a person has walked, but captures binary which activities are done or not done. This is visualised in Figure 5. The pedestrian enters at Jaarbeurs at a registered time. He then goes to the Starbucks. It is only known when he arrives and leaves Starbucks. The actual route to the Starbucks is unknown. After leaving the Starbucks the pedestrian leaves the station building via the South stairs of platform 11/12. Again the actual route towards this stairs is not known, only that and when he took the stairs.

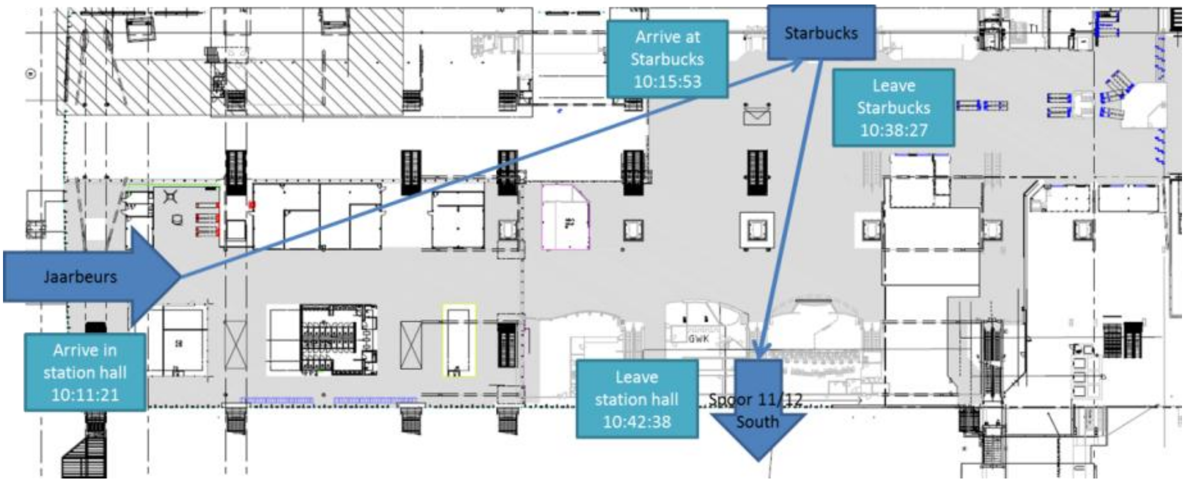

Figure 5: Example of route details measured with SMART station (at Utrecht Central Station)

This level of detail means that the behaviour at the operational level (the next step, interaction with other passengers) cannot be captured. This means that this research is aimed at higher level of detail: the tactical and strategic level of decision making (Hoogendoorn et al., 2001). See Chapter 2 for more discussion on the level of decision making.

DISCRETE CHOICE ANALYSIS

The method used to analyse the data collected with SMART station is discrete choice analysis. The decision rule used for modelling this is utility maximisation. The models used to capture these discrete choices belong to the logit family. The models considered are the multinomial logit model, the nested logit model, the cross-nested logit model and the mixed logit model. See Chapter 4 for more discussion on the discrete choice analysis.

Route choice and activity location choice have a central role in this research. Therefore two models are developed: a route choice model and a activity location choice model. A case in which the separation of these choices (related to route choice) is valid is where no activities are performed and thus no location needs to be chosen. This is the case when people head straight to the platform after entering the station building. Therefore the route choice model focuses on these departing pedestrians and what reasons they have to choose certain routes. In the activity location choice model the focus is on the location of the activity and not the route. Many routes are therefore included in the model. The focus is on departing passengers choosing one activity that can be executed on multiple locations and the reasons for them to choose a certain location.

1.3. CONTRIBUTIONS OF THIS RESEARCH

This research provides several contributions to both practice and science. These are discussed below.

On a scientific level this research contributes to a better theoretical understanding of the choice behaviour of departing pedestrians in train stations. Theories that are developed are combined into the theoretical framework of this research. The framework captures the choice behaviour of both familiar and unfamiliar passengers with the station before the trip and during the trip. It therefore shows the relationships between the choices pedestrians make and the factors and processes that influence these choices.

Relationships identified between factors and processes and route and activity location choice are selected for quantitative research when no knowledge is available or when the relationship was not quantified in the state-of-the-art. This means that new insights are gained with respect to these relations, which provide an added value to science.

On a practical level this research contributes to understanding the possibilities and applications possible of the data collection method SMART station. Its suitability for researching pedestrian choice behaviour is assessed.

The findings and insights related to the factors and processes that influence the choice behaviour of departing pedestrians in the station building can help in the development of a management strategy. These findings can also be used for future adaptions to the station building.

1.4. THESIS OUTLINE

The research approach is visualised in Figure 6. The research is split into two parts, theory development and theory assessment.

The theory development starts with literature on the choice behaviour of departing pedestrians in train stations. This is divided into three parts: the choices made in the train station, the factors that influence the choices made in the train station and the relationships amongst the choices made in the train station. This leads to the theoretical framework.

The theoretical framework is too large to be fully included in the quantitative research. Therefore a selection in the relationships identified must be made. This is based on the availability of knowledge in the state-of-the-art and whether the relationship can be captured using SMART station. The SMART station method is therefore tested on its suitability.

For the theory assessment, the data from SMART station need to be prepared in order to be useful in the model estimation. The prepared data is then validated. The validated dataset provides input for the model estimation of the two models on route choice and activity location choice. The literature on the discrete choice analysis also provides input into the models.

The estimated models then provide feedback to the created theoretical framework. This feedback is used in the conclusions and recommendations. The model estimation also provides input into the conclusions and recommendations.

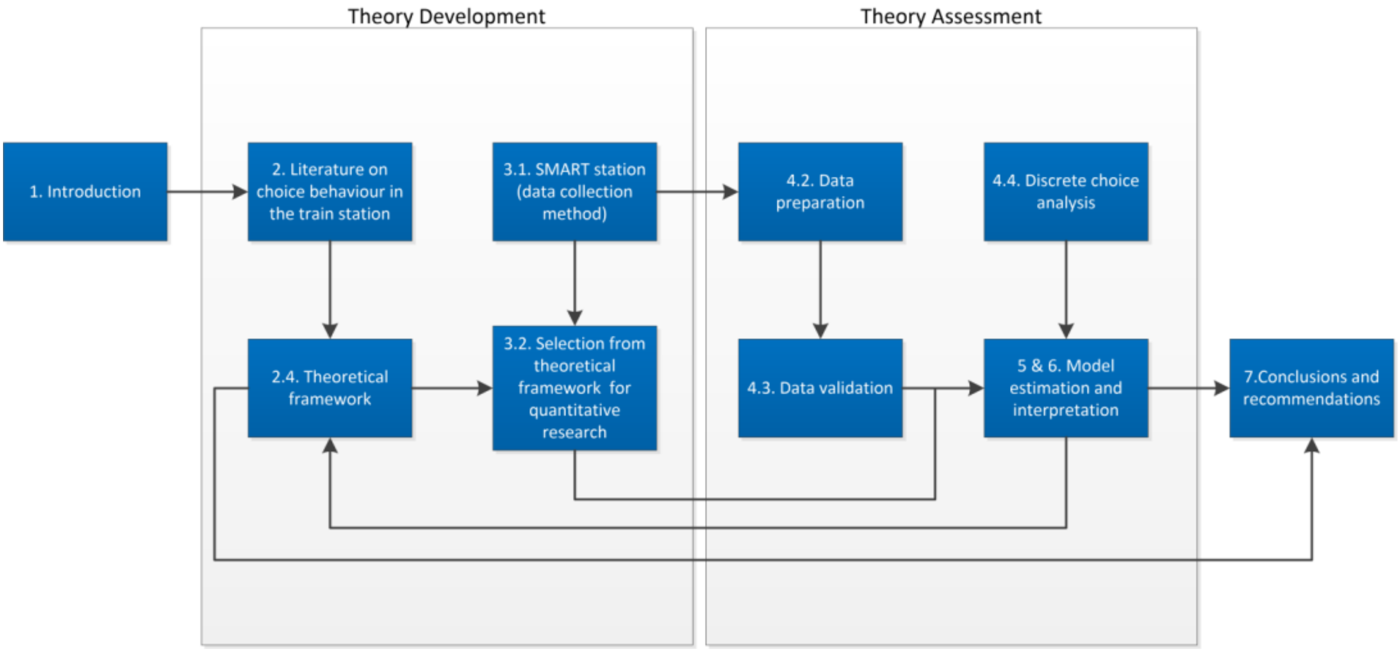

Figure 6: Research approach

2. LITERATURE REVIEW ON CHOICE BEHAVIOUR OF DEPARTING PEDESTRIANS IN TRAIN STATIONS

Understanding choice behaviour of pedestrians in train stations, starts with knowing the choices that are made in the train station. This is addressed in Section 2.1. The second part of understanding their choice behaviour is knowing which factors and processes influence the choices made in the train station (Section 2.2). The following question is discussed in this section:

‘(1) Which factors and processes influence the route and activity location choice behaviour of departing pedestrians within train stations?’

The third part of understanding the choice behaviour of pedestrians in train stations lies in knowing which relationships are present between the choices to be made in the train station. This is discussed in Section 2.3. The following question is discussed:

‘(2) Which relationships exist between route and activity location choice behaviour of departing pedestrians within train stations according to the state-of-the-art?’

When these aspects are all known, the theoretical framework can be created. This is discussed in Section 2.4. The chapter is finalised with conclusions on the choice behaviour of departing pedestrians.

2.1. CHOICES OF DEPARTING PEDESTRIANS IN TRAIN STATIONS

As mentioned before several choices are made by departing travellers in the train station. The most obvious choice made in the train station is activity choice (what to do in the station). Activities are usually used as the purpose for travel. Activity choice, however happens on different levels. On the higher level the *main activity* or the purpose for travel is chosen. This can be e.g. visiting the city Maastricht. On the lower level is the *activity choice while travelling* towards the main activity is chosen. This can be e.g. visiting the AH to Go in the station building while travelling to the city Maastricht. In the dictionary an activity is defined as a thing a person or group does or has done (Oxford Dictionaries, 2013b). If one would ask a person what he did yesterday he would say that he visited Maastricht and would never mention visiting the AH to Go. The main activity is thus more remembered than the activity while travelling.

Activity choice in literature often addresses the main activity, for example in geography. Less attention is paid to activity choice while traveling in the state-of-the-art. It is often assumed to be determined in advance of the trip and therefore less interesting for research. This research focuses on *activity choice while traveling*. An activity in this case is for example buying a coffee. The definition of activity choice used in this study is the following:

‘Activity choice is the selection of activities a departing passenger performs in the station building’

Related to activity choice three more choices are identified. Activities need to have a relative importance compared to other activities (activity hierarchy). This is necessary because the set of potential activities can be much larger than the time available to a person. It needs to be decided which activities are important and which are less important. Next to that, the set of activities needs to have a sequence in order to be performed (activity sequence). Finally, the activity often can be executed at different locations (activity location choice).

Priority needs to be allocated to each activity. This results in an activity hierarchy. This is partially dependent on the nature of the activity: mandatory or optional (Hoogendoorn & Bovy, 2004). An example of a mandatory activity within the train station is buying a ticket. An example of an optional activity is buying a cup of coffee. The hierarchy that is assigned to activities only accounts for a certain set of activities. Other situations or with other activities can result in a different priority assigned. The definition of activity hierarchy in this research is the following:

'The activity hierarchy is the relative priority assigned to the activities in the activity set of a departing passenger in the station building'

The set of activities is provided a sequence by the departing passenger performing them. There is a certain time budget, priority and duration for each activity and the complete set of activities (Root & Recker, 1981; Dijkstra et al., 2009). The sequence can be adapted while performing it (Root & Recker, 1981). This can be necessary e.g. when the time budget is smaller than the time needed for performing the activities. The definition of the activity sequence in this research is the following:

'The activity sequence is the order in which a departing passenger performs all activities of the activity set in the station building'

Hoogendoorn & Bovy (2004) state that activity scheduling has not been comprehensively studied in the past. Furthermore, they claim that there are indications that pedestrians optimize the order of activities based on utility. Ettema et al. (1993), on the other hand, state that people are not able to find the best sequence from many possibilities because the effort for scheduling is too high. In their model the expected utility is weighted against the extra effort it takes to find a new sequence. These studies show that multiple theories are present about the formation of activity sequences.

For all activities in the activity set it needs to be determined at which locations these activities will be performed. Each activity can often be performed at different locations. An example of an optional activity is buying a cup of coffee. This can be done at the Starbucks but also at the Broodzaak. These locations differ geographically. One can be more convenient due to the final destination of the traveller (the train). It can, for example also be the case that a person prefers the coffee from the Starbucks over the coffee from the Broodzaak and regardless of the convenience of the location, chooses that way. An example of a mandatory activity is buying a train ticket. This can be done at the ticket vending machine or at the tickets and service shop. The choice can be made based on convenience. However, the tickets and service shop will attract more people that are unknown to train travel in the Netherlands. The definition of activity location choice in this study is the following:

'Activity location choice is the selection of the location at which each activity of the activity set is performed by a departing passenger in the station building'

Next to the choices related to activity choice a departing pedestrian also chooses a route. A route is defined in the dictionary as 'a way or road taken from a starting point to a destination' (Oxford Dictionaries, 2013a). The definition of a route is slightly different in this research. Mainly, because the route connects all the activities performed. Therefore the definition of a route is the path connecting all activities performed in the train station. As mentioned before, in this research this is expressed in a binary sequence: one binary number for each activity.

Route choice is possible when multiple routes are available to a traveller. This happens on a high level, when travelling to the main activity or on a low level, when performing activities while travelling. The focus of this study is on the station building and therefore on the low level. For both practical and scientific purposes it is interesting to know how people are divided over routes and how they choose these routes (both on the high and low level). Issues that have received attention in the state-of-the-art are e.g. modelling route choice behaviour (e.g. Antonini et al., 2006 or Borgers & Timmermans, 1986), identifying factors that influence route choice behaviour (e.g. Seneviratne & Morall, 1985 or Verlander & Heydecker, 1997) and congestion or crowd formation (e.g. Helbing et al., 2001 or Cheung & Lam, 1998). The definition of route choice in this study is the following:

'Route choice is the selection of the route connecting all activities performed in the station building by the departing passenger'

In conclusion, five choices are identified that departing pedestrians make in the station building. These choices are activity choice, activity hierarchy, activity sequence, activity location choice and route choice.

2.2. INFLUENCES ON THE CHOICE BEHAVIOUR OF DEPARTING PEDESTRIANS

In this section factors and processes are identified that influence the choice behaviour of departing pedestrians in train stations. In this identification process a selection is made on the extent to which these factors and processes influence the choices concerned and what their relevance is for the train station environment. The selection is based on the state-of-the-art on choice behaviour and on observations from train stations. Appendix A provides a list of all factors and processes identified before this selection.

According to Daamen (2004) factors influencing route choice can be divided into four groups: network characteristics, route characteristics, personal characteristics and trip characteristics. In this research a different grouping structure is used. The network, route and trip are not as important for activity location choice as they are for route choice. Therefore, these are combined into system factors. The main business in the station environment is the train. Therefore, public transport factors are included. The personal factors identified by Daamen (2004) can also be used in this research. Another group that that affects both activity location and route choice in train stations are the external factors. The external factors cannot be influenced but it is important to know which 'forces' are present in the station building. In summary, the groups identified for this research are personal factors, system factors, public transport factors and external factors.

PERSONAL FACTORS

The personal factors that influence choice behaviour can be divided into three groups. These are general factors of each passenger, the factors that belong to the trip they make at that point, and the learning process of the passengers. In Table 1 a schematic overview of these factors and processes is provided.

Table 1: Schematic overview of the personal factors that influence choice behaviour within the train station

Personal factors	General personal factors	Age
		Gender
		Orientation
		Familiarity
		Emotional state
		Personal uncertainty
	Trip factors	Group composition
		Trip purpose
		Time spent in the station
		Time of day or week
		Impulse behaviour
	Learning processes	Cognitive learning
		Habit
		Choice inertia

General personal factors

The general personal factors concerned are age, gender, orientation, familiarity, emotional state and the level of personal uncertainty.

- ***Age*** → Extensive literature is available for this factor. Hill (1982) and Seneviratne & Morall (1985) for example addressed this factor. They concluded that seniors (60+) choose routes based on different reasons than other age groups. Seniors place a higher value to the safety of the route (Seneviratne & Morrall, 1985). Children, on the other hand choose differently. They prefer routes with a higher complexity compared to other age groups (Hill, 1982). The complexity relates to the number of turns one needs to make. It is plausible that in the station environment people behave similar to these findings. However, children are usually not on their own in the station building. Therefore, this does not hold for children.

- ***Gender*** → Extensive literature is available on this factor. Hill (1982) and Seneviratne & Morall (1985) addressed gender. They found that the reasons for choosing a route do not differ between men and women (Seneviratne & Morrall, 1985). However, women tend to choose more complex routes than men (Hill, 1982). This difference stems from the research method used (Golledge, 1999). Seneviratne & Morall (1985) asked people about reasons for choosing a route (stated preference). While Hill (1982) observed the routes of people (revealed preference). It is again plausible that behaviour in the station environment is similar to these findings.

- ***Orientation*** → The orientation of a pedestrian relates to the preference for left or right. This is strongly dependent on the environment in which a person grew up. Individuals in countries where one drives/walks on the left side of the road (e.g. Great Britain or India) have a stronger preference towards the left side and thus activities that are located on the left. Whereas pedestrians in countries where one drives/walks on the right, will have a stronger preference towards the right side and activities on located on the right. Other factors of influence are attractiveness of the alternative and crowdedness (Bishop et al., 2001).

- ***Familiarity*** → For this factor extensive literature is available. This was addressed by e.g. Dijkstra et al. (2009) and Bovy & Stern (1990). People tend to choose routes and activities they know and are reluctant to choose new ones (Golledge, 1999). Daamen & Hoogendoorn (2003) distinguish between two groups of pedestrians; commuters and tourists. The commuters are familiar with the station environment, because they go there nearly every day. Tourists, on the other hand, are unfamiliar with the station. Van Hagen (2011) acknowledges this. He introduces two terms for familiar and unfamiliar passengers; must (familiar, commuter) and lust (unfamiliar, social-recreational) travellers. All studies found that familiar travellers behave significantly different from unfamiliar travellers. Unfamiliar travellers show more searching behaviour and do not always make the ‘logical’ choices because their known choice set is smaller than of familiar travellers (Hoogendoorn-Lanser, 2005). Therefore, this factor is important to take into account in this research.

- ***Emotional state*** → This was addressed by e.g. Dijkstra et al. (2009). It depends on e.g. the crowdedness in a person’s surroundings (Lee et al., 2001). This can relate to frustration, panic and feeling of safety. Emotional state however is a subjective factor that is hard to capture. Therefore it is only described in the state-of-the-art in a qualitative way.

- ***Level of personal uncertainty*** → This factor relates to the emotional state. It is related to the level of control that is required by a person (van Hagen, 2011). Uncertainty can be expressed in the following ways:

- As *familiarity*: a lust passenger (social-recreational) often feels the need to control the situation (NS Stations & Altuïtion, 2013), but due to their unfamiliarity uncertainty arises.
- As *density factor:* people can feel restricted in their movement and available space which can lead to stress and uncertainty (Stokols, 1972).
- As *time factor*: the time margin needed for people to catch the train. It was observed that people arrive at different times before train departure. They therefore have different perspectives on the risk of missing the train.
- As *planning factor*: can all activities that are planned be performed before the train departs (Hoogendoorn & Bovy, 2004)?

In the station building the familiarity related to uncertainty plays an important role when looking at activity location choice. One cannot perform an activity on a location is not familiar with. Also the time factor plays a large role. However as the frequency of the trains increases this time factor reduces, because the penalty of missing the train is lower.

Trip factors

The trip factors concerned are trip purpose, time of day or week, arrival time at the station, impulse behaviour and group composition.

- ***Trip purpose*** → This factor was addressed in several studies, e.g. Golledge (1999) and NS Stations (2013). Several trip purposes are distinguished. The most common classification is: work, school, business and social-recreational. It was found that a significant difference arises in choice behaviour when looking at trip purpose (e.g. van Hagen, 2011). This is visible in the routes chosen: people heading for work or school pay less attention to attractiveness of the route (Seneviratne & Morall, 1985).

- ***Time of day or week*** → Golledge (1999) and Seneviratne & Morall (1985) discussed this factor. The choices and reasons for choosing a route can change during the day or week (Golledge, 1999). This factor also relates to the trip purpose of a person. People travelling at 8.00 AM on a weekday are mostly going to work, while people travelling on Sunday afternoon are mainly social-recreational. However, not much research considering this factor has been carried out and none has been quantified. This offers opportunities for this research.

- ***Time spent in the station*** → The time spent in the station building related to the departure time of the train differs largely amongst passengers. Some people arrive ten minutes in advance while others arrive only two minutes in advance. This has to do with the risk people accept of missing the train and the impact this has on their waiting/travel time. Risk prone people do not mind much if they miss the train, while risk averse people make sure they do not miss the train. This depends on the time table (see public transport factors). Next to that, the time spent in the station also relates to the activities performed. More time available means that more activities can be performed, on the other hand people can also come earlier to perform the activities. This was never quantified in research and provides an interesting opportunity for this research.

- ***Impulse behaviour*** → This behaviour was researched by Dijkstra et al. (2009). It can arise while walking through the station building, e.g. in noticing the Broodzaak and suddenly feeling the urge to buy a sandwich. This behaviour was not intended beforehand. Therefore it influences also the decisions that a pedestrian makes afterwards.

- ***Group composition*** → This factor was researched by e.g. Hill (1982) and Bierlaire & Robin (2009). Actual choices can differ from being alone to being in a group. Also the role a person has within the group influences the choices (leader or follower).

Learning processes

The learning processes concerned are cognitive learning, habit and choice inertia.

- ***Cognitive learning*** → Bovy & Stern (1990) addressed this process. People learn from an experience for future situations. They form a mental map that consists of explicit, geometric, cartographic and artistic representations of reality. Passengers differ in the way and speed they learn.

- ***Habit*** → This was addressed in multiple researches. Bogers (2009) tried to capture habit in her research, by making people repeatedly answer a questionnaire about their route choice. This way the learning curve of people could be retrieved. Hoogendoorn & Bovy (2004), Hill (1982) and Seneviratne & Morall (1985) researched this process. A habit arises when people repeat the same pattern every time without 'thinking'. People like to retrace their steps over learning new things, because it takes less effort (Golledge, 1999). Golledge (1999) also states that a person creates a habit within +/- 6 visits.

- ***Choice inertia*** → It takes a while to adapt to a new situation: for example, when new routes or activities are presented. This is known as choice inertia. This process was addressed by Bovy & Stern (1990). However, it is very difficult to capture choice inertia due to the long term observation of individuals needed to deduce it.

SYSTEM FACTORS

The system factors that influence choice behaviour of pedestrians concern the station building. Within this group a distinction is made between location factors and route attributes. This is visualised in Table 2.

Table 2: Schematic overview of the system factors that influence choice behaviour in the train station

System Factors	Location factors	Location dimensions
		Route constraints
		Visibility
		Waiting area characteristics
		Information
		System uncertainty
		Amount of shops or activities
	Route Attributes	Travel time
		Distance
		Crowdedness
		Number of attractions
		Comfort
		Weather protection
		Directness

Location factors

The location factors concerned are dimensions of the location, route constraints, visibility, waiting area characteristics, information, system uncertainty and the amount of shops and other activities available.

- ***Dimensions of the location*** → Lots of research addressed this factor, e.g. Timmermans (2009), Schadschneider et al. (2008) and Bierlaire & Robin (2009). The dimensions of an area influence the movement of people and have an impact on their route and location choice. Next to that, it influences the flow through the station building. If a station is logically designed the flow will go fluently.

- ***Route constraints*** → This factor was discussed by Bovy & Stern (1990). Some routes or locations are not accessible by all travellers. These constraints can be personal or systematic. An example of a personal constraint is a wheelchair that cannot be used on the stairs. An example of a systematic constraint is cyclists that are not allowed in the station building.

- ***Visibility*** → The visibility relates to the visibility of the route inside the station building. When a person is does not know the station building routes that are partially invisible, they are easily skipped. The location of the invisible route is also often misjudged (Montello, 1991). Because one sees the visible locations and routes first, one is more tempted to choose those. This factor therefore also relates to the familiarity of the individual with the station building.

- ***Waiting area characteristics*** → This factor was addressed by Van Hagen (2011). A pleasant waiting area increases the experience of people. The waiting area can be a platform or a place in the station building. Where people wait might influence the location of their activities and its location choice. This is addressed in practice by adapting the waiting areas into more pleasant areas.

- ***Information*** →The information provided to people, both static and dynamic, influences the choice behaviour. This was found by e.g. Van Hagen (2011). Travellers like to control the situation and information provides them with this opportunity (related to train departure times, routing etc).

- ***System uncertainty*** → This relates to the uncertainty that arises due to the system. It can be expressed in different ways;
    - As *information provision:* the search for signs and train departure information. If this information is not found or no satisfactory information is found this leads to uncertainty. The station lay-out can help people orientate quickly and prevent confusion (Nie, 2000).
    - As *train operations*: delays and changes in the timetable. Taylor (1994) found a negative relationship between delay and uncertainty at airports. The same is plausible at train stations. Knowing about the delay or change is usually perceived as more positive than not knowing (van Hagen, 2011).

- ***Amount and type of shops***→ The availability of multiple activities or different types of activities influences choice behaviour (especially related to activity choice). It was observed that the presence of some activities at one station triggered different behaviour than when those activities were not present. An example is the presence of Starbucks at Utrecht Central station. This activity triggers people to stay for a longer time. The opposite is the AH to Go where one 'grabs a coffee on the go'.

Route attributes

The route attributes concerned are travel time, distance, crowdedness, number of attractions, comfort, weather protection and directness.

- ***Travel time*** → This attribute was addressed in many studies, e.g. Cheung & Lam (1998) and Daamen & Hoogendoorn (2003). This factor is often assumed to be the critical attribute for choosing a route. In the train station heavy queuing can occur locally. This significantly influences the travel time in some cases. Pedestrians can therefore switch routes if too long a queue appears (Voskamp, 2012). However travel time is only researched quantitatively in the train station at places where a lot of pedestrians are present at once (train alighting). This has not yet been researched for departing passengers, which offers an interesting opportunity for this research.

- ***Distance*** → This attribute was discussed in many studies, e.g. Seneviratne & Morall (1985) and Borgers & Timmermans (1986). According to Seneviratne & Morrall (1985) routes that are shortest in distance versus shortest in time are ambiguous for pedestrians. Travellers do not know whether they optimise for time or distance. However, some studies e.g. Borgers & Timmermans (1986) use distance in their research. The distance was often addressed in literature, but only related to route choice. However, it was observed that people also take distance into account when choosing a location for their activity. If you have a choice between two equal services the closest is often selected, unless there are specific preferences for the other service. This has not been researched before and provides an interesting research opportunity.

- ***Crowdedness*** → Many studies addressed this attribute, e.g. Lee et al. (2001), Helbing (1997) and Daamen & Hoogendoorn (2003). Crowdedness arises when the demand for space is larger than the available space (Stokols, 1972). The Netherlands Railways (NS) translate crowdedness to the level of service offered to travellers (de Vos, 2011). This level of service is based on Fruin (1971). When the level of service is too low or the crowdedness to high, a high travel time is expected. Therefore route changing can occur (Voskamp, 2012). When it is too crowded the level of comfort drops (Lee et al, 2001).

- ***Number of attractions*** → This attribute was discussed by e.g. Hoogendoorn & Bovy (2004). Seneviratne & Morall (1985) state that the importance of this attribute is related to the trip purpose. A person travelling to work is less triggered by attractions on the route than a person visiting a city. This can be translated again to the difference between must and lust passengers (van Hagen, 2011).

- ***Comfort*** → This factor is very subjective and differs between modes (Bovy & Stern, 1990). In a car, comfort is defined differently than while walking. In the train station, comfort translates for example to the use of stairs versus escalator (e.g. Lee et al., 2001). In general, the escalator is experienced as more comfortable.

- ***Weather protection*** → This can relate to the level of comfort travellers experience. If a route offers protection against the weather (especially during rain) it was observed this was more often chosen than a route which exposed people to the weather. On the other hand, it was observed that (up to a certain limit) in sunny weather people tend to choose routes which expose them to this weather. On the other hand, if the sun is too fierce people try to avoid these routes.

- ***Directness*** → This attribute was addressed by Helbing (1997). A route is classified as direct when a pedestrian can walk in a straight line to his or her destination. The possibility of hindrance arises in the shape of obstacles or other pedestrians. The directness is therefore also defined in relationship to visibility and number of turns in a route.

PUBLIC TRANSPORT FACTORS

The public transport factors that influence choice behaviour of pedestrians relate to the station environment and network. The public transport factors concerned are time table (also related to for example frequency and train service type) and train operations. Table 3 provides an overview of these factors.

Table 3: Schematic overview of public transport factors that influence choice behaviour in the train station

Public Transport Factors	Time table
	Train operations

- ***Time table*** → The time table relates to the planning of trains. This covers the platform that is addressed and times at which trains depart. But also the train service used by a pedestrian to leave the station and the frequency of the trains. The train services present on the train network in the Netherlands are sprinter trains (local), intercity trains and international trains. This train service is also reflected in the frequency. Local trains have a much higher frequency than international trains. This also reflects the attitude of pedestrians towards risk. A train service with a lower frequency means a longer waiting time if missed, therefore the risk aversion is expected to be higher with a lower frequency.

  In the state-of-the-art the time table as stated here is not addressed and therefore provides interesting opportunities for this research. Research on e.g. the influence of dwell times in stations on the capacity of the network (Parkinson & Fisher, 1996) or influence of train design on station capacity (Vuchic, 2007) do relate to this factor. Parkinson & Fisher (1996) define the dwell time as the number of people that want to board and alight a train. This was also researched by Qi et al. (2008). It was observed at Schiphol station that more people boarding and alighting increases the dwell time.

- ***Train operations*** → The train operations are defined as the execution of the planned time table. This can relate to the trains departing on time or with a delay, trains switching platforms or not going at all. It was observed that delays in the train service influence travellers in e.g. the activities they perform (only when enough time is present). Van Hagen (2011) addressed this with respect to waiting behaviour of travellers. The train operations as defined here, are however not researched in the state-of-the-art and therefore provide an opportunity for this research.

EXTERNAL FACTORS

The external factors that influence choice behaviour of pedestrians are related to factors that cannot be changed by people and do not belong to the system or public transport factors. The external factors concerned are environment and weather. Table 4 shows an overview of these factors.

Table 4: Schematic overview of external factors that influence choice behaviour within the train station

External Factors	Environment
	Weather

- ***Environment*** → This factor was addressed by Bovy & Stern (1990). They state that the actual choice a traveller makes strongly depends on the environment in which they are moving. Environment can be defined in multiple ways. It can be seen as the station building and its dimensions. But it can also be seen on a larger scale in which it relates to different cities or countries. In different cities or countries different cultures and standards are present.

- ***Weather*** → This was researched by SBB in Switzerland (Thurau, 2013). Stations that are completely indoor, like Schiphol station, are not affected by this factor. For other stations they found that temperature has a higher influence on the waiting location of passengers than rain. Passengers searched for sheltered routes when it was very warm. The relationship for rain was weaker.

In this section all factors and processes that influence the choice behaviour of departing passengers in train stations are identified. These relationships are summarised in Table 5. If the relationship is not mentioned in the table, this means that it was not found in pedestrian literature. It could be unknown (not researched) or it could be present in other areas than the pedestrian literature (like social geography) that were not researched for this study.

Table 5: Relationships between influence factors and choice aspects

	Factors and processes	Choice of activities	Activity hierarchy	Activity sequence	Location choice	Route choice
General personal factors	Age					x
	Gender					x
	Orientation				x	x
	Familiarity			x	x	x
	Emotional state		x	x		x
	Personal uncertainty			x	x	x
Trip factors	Group composition	x		x		x
	Trip purpose	x	x	x		x
	Time spent in the station			x	x	x
	Time of day or week	x			x	x
	Impulse behaviour	x		x		x
Learning processes	Cognitive learning			x	x	x
	Habit			x	x	x
	Choice inertia			x	x	x
Location factors	Location dimensions				x	x
	Route constraints					x
	Visibility					x
	Waiting area characteristics			x	x	
	Information			x	x	x
	System uncertainty			x	x	x
	Amount of shops	x		x	x	
Route attributes	Travel time				x	x
	Distance				x	x
	Crowdedness				x	x
	Number of attractions					x
	Comfort					x
	Weather protection					x
	Directness					x
Public transport factors	Time table			x	x	x
	Train operations			x	x	x
External factors	Environment	x				
	Weather				x	x

2.3. RELATIONSHIPS BETWEEN ROUTE AND ACTIVITY LOCATION CHOICE

The previous section identified the factors and processes that influence the choices made in the train station. In this section the relationships between the choices made in the train station are reviewed by means of the state-of-the-art.

Root & Recker (1981) stated that it is essential to understand how various factors influence the selection of activity sites, timing of activities and creating a sequence of activities for an individual. They state that in order to fully understand this, a theoretical framework capturing the relations, needs to be created. They introduced their first ideas in the framework shown in Figure 7. In this framework two phases are identified: pre-travel and travel. In the pre-travel phase the activities are chosen and an activity program or sequence is planned. The decisions in this phase influence the choices made during the travel phase. The activity sequence can be adapted when the situation changes. According to them a comparison is made between e.g. time (needed and away from home), location and the remaining activities. This decides whether the activity sequence can be fulfilled or needs adaption. A feedback relationship from the activity sequence to the planning of the activity sequence (pre-travel) and the execution of the next activity in the sequence (during travel) is made. This provides information about the possibilities to execute the activity sequence.

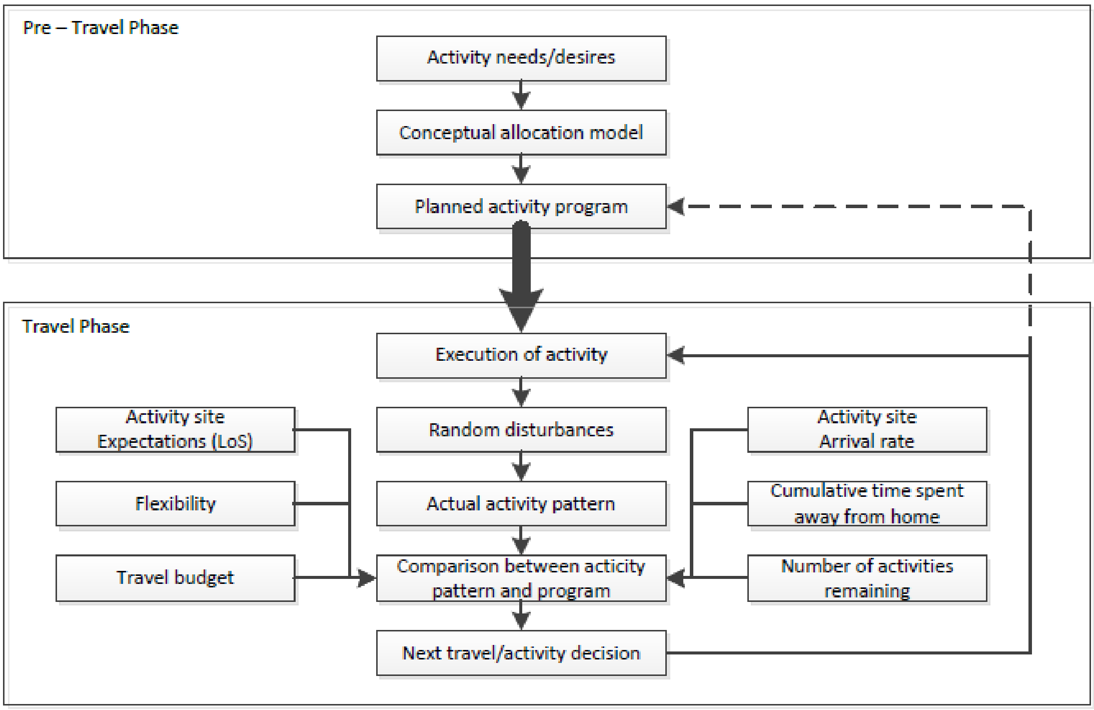

Figure 7: Theoretical framework on activity choice and sequences (Root & Recker, 1981)

This framework provides interesting relationships that can be applied to the situation in the train station. The planning of the activity sequence before travelling also applies to the train station. The adaption of the activity schedule during the trip due to e.g. time constraints is suitable for the situation in the train station. When less time is available (e.g. one arrives later and the train departs on time) adaptions to the planned schedule need to be made. The location of the activity that influences the resulting sequence also provides an interesting insight. Finally, the addition of the feedback relationship is applicable to departing passengers in the train station.

Hoogendoorn et al. (2001) introduce a framework that shows different levels of pedestrian behaviour (see Figure 8). These levels are comparable to the levels introduced by Root & Recker (1981). Hoogendoorn et al. (2001) only distinguish three levels instead of two: strategic (pre-travel phase), tactical (travel phase) and operational (travel phase). Daamen (2004) adapted this framework for her research.

According to Hoogendoorn et al. (2001) activity choice is done on a strategic level. Route choice, location choice and the activity schedule are then determined on a tactical level, thus during the trip. The tactical and strategic level influence each other. The operational level relates to for example walking or interaction with other people. This is basically the next ‘step’, for example to move side ways to avoid a collision with another pedestrian. This is too detailed for this research. It is not possible to capture these microscopic movements by means of the data collection method SMART station. This study, therefore focuses on the strategic and tactical level.

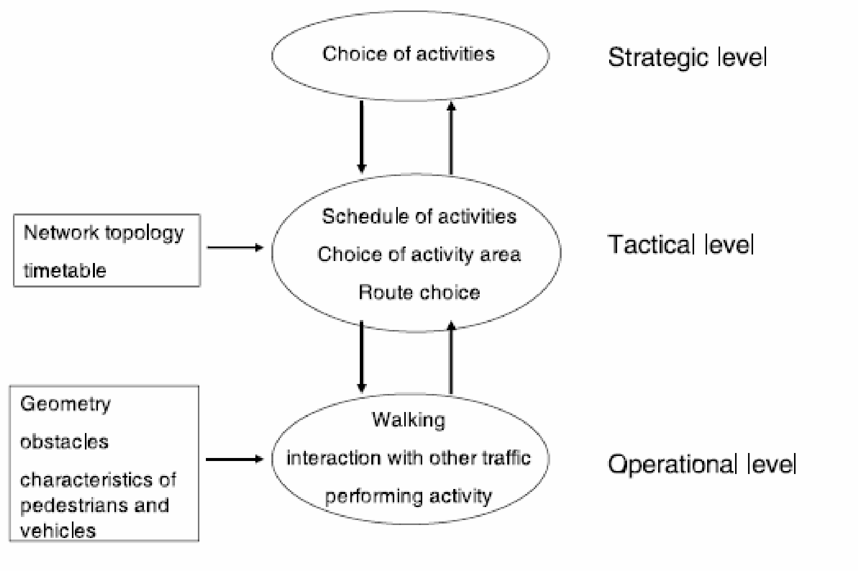

Figure 8: Levels in Pedestrian Behaviour (Hoogendoorn et al., 2001)

All choices on the tactical level relate to one another. Hoogendoorn & Bovy (2004) assume that pedestrians choose the activity location and route simultaneously. They use the concept of utility in their research. Pedestrians try to maximize their utility by optimising the route and activity location(s). With respect to the activity schedule, they assume that it is fixed. All feasible activity schedules are listed and the optimal one is chosen. As mentioned before, no consensus exists in literature about how people determine the activity schedule (Ettema et al., 1993; Hoogendoorn & Bovy, 2004).

The framework of Hoogendoorn et al. (2001) offers interesting insights for this research. The location choice of an activity is decided on while travelling, together with the route choice. This can be applied to departing passengers in the train station. Also the sequential decision on route choice and location choice seems applicable.

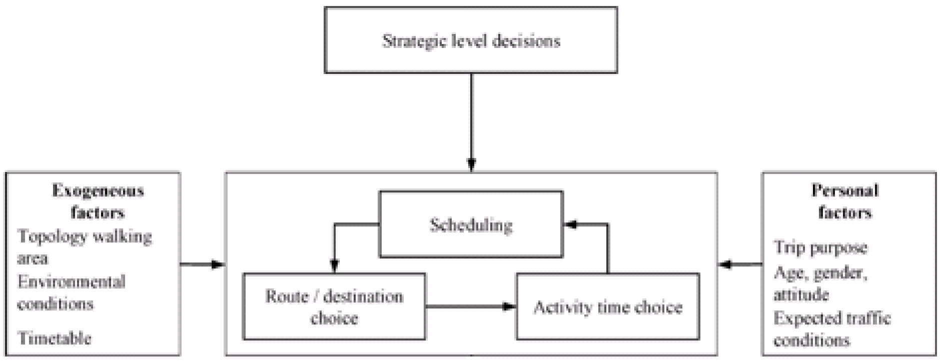

Figure 9: Route choice process for the individual traveller (Hoogendoorn & Bovy, 2005)

Another research performed by Hoogendoorn & Bovy (2005) uses similar relationships between the aspects of activity choice and route choice (see Figure 9). They also introduce a simultaneous choice process between route choice and activity location choice. There are however some differences in their framework compared to the framework of Hoogendoorn et al. (2001). The strategic level decisions (activity choice) influence the decisions made on a tactical level but not the other way around. In Hoogendoorn et al. (2001) the strategic level decisions were also influenced by decisions on a tactical level (feedback relation). The experience gained during that trip is taken into account for the next trip. The presence of this feedback relationships seems plausible in the station area. Experience can lead to adapting or changing the decisions made on a strategic level. Hoogendoorn & Bovy (2005) also introduce a relationship between activity scheduling and the route and activity location choice.

The frameworks introduced in this section can be used for the creation of the theoretical framework for this study. Each of the frameworks has provided interesting insights in choice behaviour. The relationships identified by Root & Recker (1981) provide useful information on the activity sequence and activity location choice in the pre-travel and travel phase. The distribution of choices over the strategic and tactical level of Hoogendoorn et al. (2001) offer the basis for the framework in this research. Finally, the relationships presented by Hoogendoorn & Bovy (2005) are also (partially) used for the identification of the framework.

2.4. THEORETICAL FRAMEWORK

In Section 2.2 the issue arose that familiar people behave significantly differently in the station area from unfamiliar people. This was stated in multiple research. This is a factor that addresses everyone, a pedestrian is either familiar or unfamiliar.

The familiar travellers visit the train station often (commuters or business). Unfamiliar travellers, on the other hand, visit the train station only on rare occasions (social-recreational). The difference between commuters and social-recreational travellers is used by NS in their pricing policy. People are rewarded for travelling outside the peak hours, they can receive a discount of 40% (NS, 2014). For most commuters this is not possible, because they have to be in the office in time. For social-recreational travellers this is usually not the issue and therefore these people are often present in off-peak hours. This means that NS also knows and applies the knowledge about this difference in familiarity. The familiarity also relates to different learning processes of the two groups (habit formation) (Hill, 1982).

In the choice process the familiarity translates to differences in the choices made in the pre-travel phase. Because of their knowledge of the station, the familiar passenger is able to plan more precisely what choices he is going to make. The unfamiliar passenger cannot do this to the same extent, because he lacks knowledge on the station. Because of the significant difference between familiar and unfamiliar passengers in their choice behaviour it was decided to distinguish between these groups in the theoretical framework on choice behaviour. Section 2.4.1 covers the theoretical framework for familiar passengers and section 2.4.2 addresses the theoretical framework for unfamiliar passengers.

2.4.1. THEORETICAL FRAMEWORK FOR FAMILIAR DEPARTING PASSENGERS

The familiar passenger is known to the station and visits it on a regular basis (commuter/business) (van Hagen, 2011). Golledge (1999) states that in repetitive trips a habit will show within five or six times. This means that the familiar passenger is subject to habitual behaviour. The familiar passenger will remember the trips he made before. He will take this into account when planning the new trip. If everything goes according to this plan, no adaptions are needed during the trip. However, if the situation does change decisions are made during the trip. For example, the situation can change due to a delayed train, changed platform, different traffic operations (e.g. crowdedness), impulse behaviour or a different arrival time at the station (both earlier and later). Figure 10 shows the theoretical framework for familiar passengers in the station. The framework has two phases. These phases were also introduced in frameworks of e.g. Root & Recker (1981) and Hoogendoorn et al. (2001). The two phases are before the trip and during the trip. Before the trip relates to the pre-travel phase introduced in Root & Recker (1981) or the strategic level decisions introduced by Hoogendoorn et al. (2001). During the trip equals the travel phase or the tactical level.

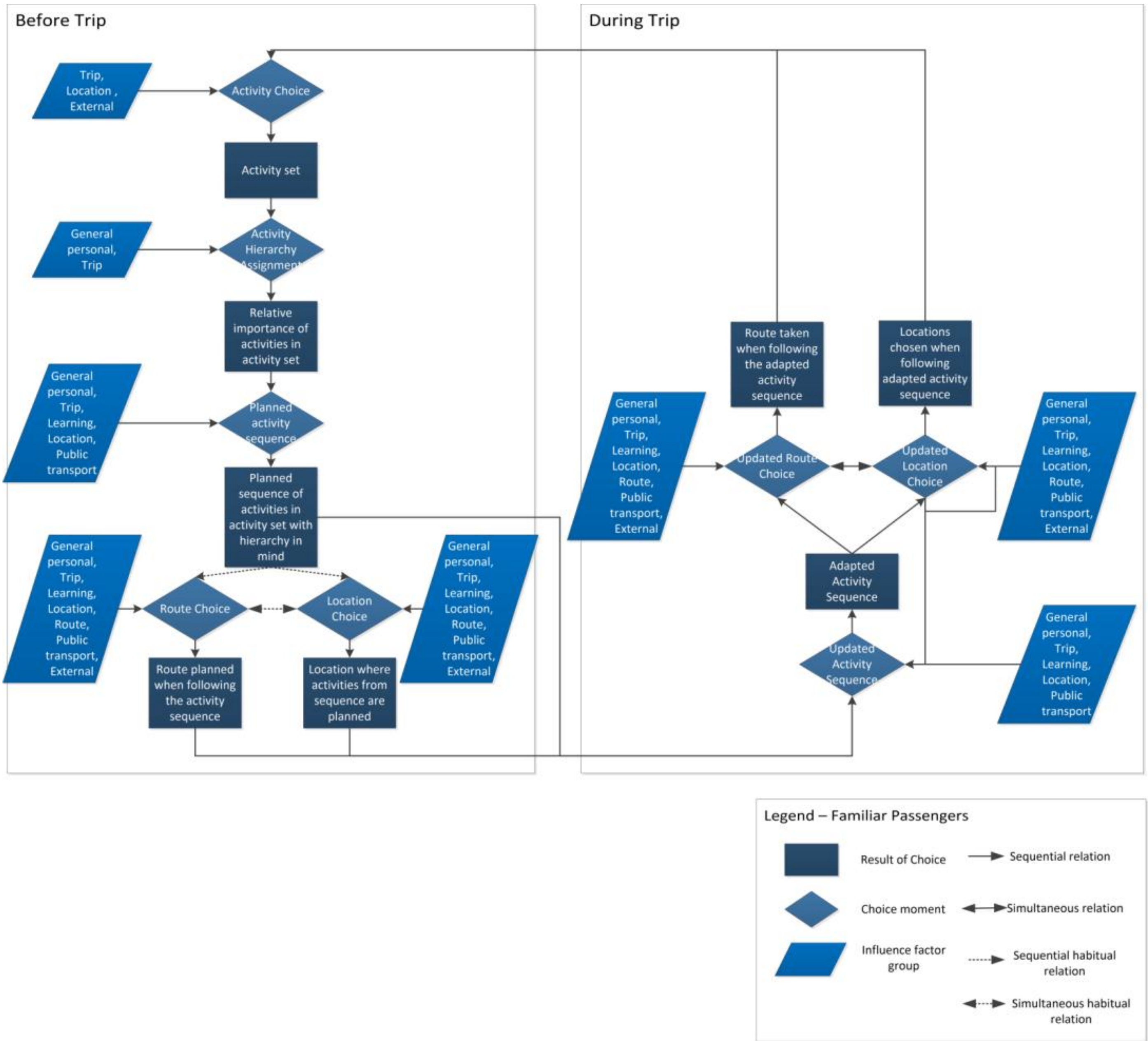

Figure 10: Framework for familiar travellers

BEFORE THE TRIP

Before the trip the decision is made on which activities to perform in the station building (e.g. Hoogendoorn et al., 2001). The input for this decision are the factors that are identified in section 2.2 which influence the activity choice behaviour of pedestrians. These are factors belonging to the trip factors, location factors and external factors. The output of the activity choice is the set of activities that is planned.

After the decision on the activity choice set, the hierarchy of the activity set is determined. This choice process is influenced by factors that are general personal and trip factors. This means that the hierarchy that is assigned to each of the activities determines the trip he is going to make. The output of this decision is the relative importance the departing passenger assigns to the activities in his choice set.

The departing passenger then decides on the activity sequence. He already knows which activities are included and what their relative importance is. This is used as input for the sequencing process. Next to that, the choice is influenced by several factors. These are general personal, trip, learning processes, location and public transport factors. This means that the sequence a passenger assigns to the activity set depends on the person itself, the learning curve of a person or their habit formation, the trip that is planned, the location of the trip, and the public transport that is visited. This evolves into a planned sequence of the activities belonging to the activity set with the relative importance in mind.

The familiar passenger is subject to habitual behaviour. This means that he uses the information gathered in previous trips and includes that in plans for the next trip. It also means that after the planning of the activity sequence this passenger already plans the route and location for each activity. An example is that you plan to buy a sandwich and a coffee. The planned route could then be entering the station building, going to the Broodzaak for a sandwich, going to the Starbucks to buy coffee and afterwards heading to platform 11/12 to board the train. Hoogendoorn & Bovy (2004) introduced the simultaneous choice for route and location. This concept is also applied in the framework for the familiar passenger. These passengers know the station well enough an are assumed to combine both choices.

The input for the planned route choice are factors that belong to all groups identified in section 2.2: the general personal factors, trip factors, learning factors, location factors, route factors, public transport factors and external factors. The output of this choice is the planned route when following the activity sequence. The input for the location choice also comes from all groups. This decision leads to the location where the activities from the activity sequence are planned.

The route choice, location choice and activity sequence provide input for the processes and choices that take place during the trip.

DURING THE TRIP

Decisions are made during the trip if the situation changes. This will first influence the activity sequence, because in the sequence the activities can be removed or changed. This means that the activity set and the activity hierarchy remain as planned. The changes in activities happen in sequence. If for example less time is available than originally planned it is possible that not all planned activities can be performed. In that case these will be deleted from the activity sequence (Hoogendoorn & Bovy, 2004). The decision for updating the activity sequence come from the planned route choice, location choice and activity sequence. Input into the activity sequence belong to the groups: general personal factors, trip factors, learning factors, location factors and public transport factors. The output from this decision is the adapted activity sequence.

Just as before the trip, the activity sequence provides input for the route choice and location choice. For these factors the same input is provided as in the phase before the trip. However, in this case the updated location choice provides input again into the activity sequence. The location of one activity influences the proceedings of the activity sequence (also seen in Roote & Recker (1981)). Also the location choice influences the location of the next activity. Most likely, the location of the next activity will be relatively close to the current activity. The output of the location choice decision is the chosen locations when following the adapted activity sequence. The output for the route choice is the route taken when following the adapted activity sequence.

The experience gathered from the performed sequence, route and locations provide input for the next trip. This serves as feedback and extra information to the familiar traveller. If for example the train was missed due to the updated sequence then next time this information can be used in the planning as to not miss the train.

2.4.2. THEORETICAL FRAMEWORK FOR UNFAMILIAR DEPARTING PASSENGERS

In this section the framework for the unfamiliar departing passenger is addressed. The unfamiliar passenger is not known to the station (van Hagen, 2011). This is due to less train travel compared to the familiar passenger. This passenger will usually not show habitual behaviour in the train station, due to his unfamiliarity (Golledge, 1999). The extent to which the unfamiliar traveller can plan his trip in advance is much less than the familiar passenger. The unfamiliar passenger cannot create a route or determine on the location of the activity. The route and location choice are therefore determined during the trip. In Figure 11 the theoretical framework for unfamiliar passengers is shown.

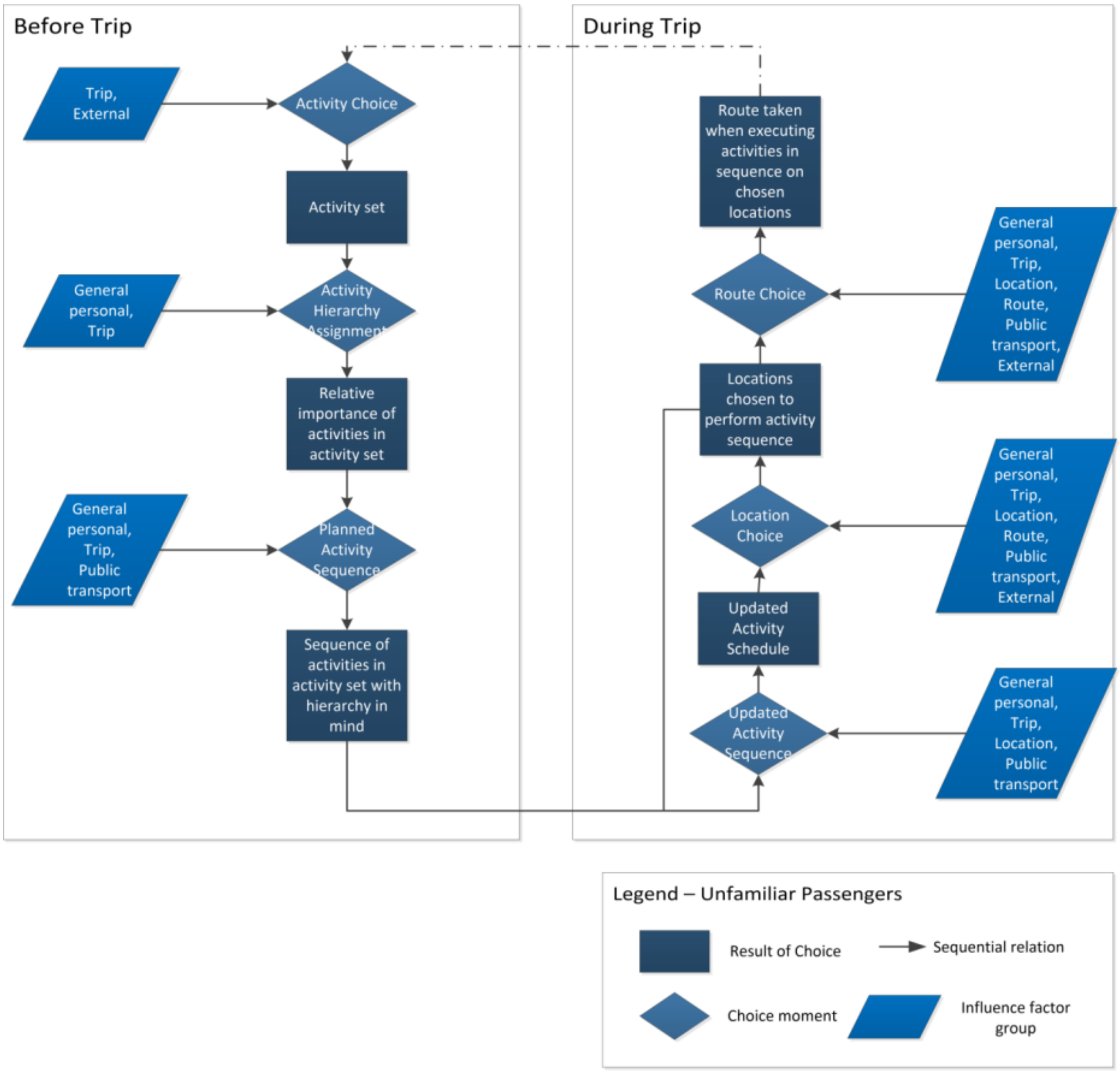

Figure 11: Theoretical framework for unfamiliar passengers

This framework is again split into two levels, just as was done in the framework for familiar passengers in section 2.4.1. Before the trip is one level and during the trip is the other level.

BEFORE THE TRIP

Before the trip the activity set is decided. The input into this choice are trip factors and external factors. This means that the activity set depends on the trip you make and the situation in which you make the trip. The output of the choice is the activity set.

After this the relative importance of each activity is assigned to the activity set. The input factors for this decision are general personal factors and trip factors. The unfamiliar passengers will probably provide the hierarchy in the same way as the familiar passenger. The output of this decision is the relative importance of each activity in the activity set.

The next step is to make a sequence of the activities planned. The unfamiliar passenger is more uncertain than the familiar passenger due to not knowing the station area. Therefore the unfamiliar passenger will try to keep control by making plans in advance (NS Stations & Altuïtion, 2013). This however costs more effort compared to the familiar passenger, who also wants to stay in control. This can lead to extra time planned for each activity, to make sure that everything can be done. The input for the activity sequence comes from the following groups: general personal factors, trip factors and public transport factors. The output is the sequence of activities in the activity set with the relative importance of activities in mind. This output serves as input for the decision made during the trip.

DURING THE TRIP

The activity sequence is updated during the trip if the situation changes or if new information is learned by visiting the station building. For this passenger type, this also translates to adding activities they were unfamiliar with. However, due to the control and uncertainty issue, this will only be done if time is available. The activity sequence gets input from the planned activity sequence and the location chosen to perform an activity. This location influences the next activities to be performed. This can for example be expressed as skipping an activity because the location is too far from the final destination (platform) and not enough time is available. The factors that influence the activity sequence are general personal, trip, location and public transport factors. The output produced by this decision is the updated activity sequence.

The next decision is the location at which the activities in the activity sequence will be performed. The factors that influence this decision belong to the following groups: general personal factors, trip factors, location factors, route attributes, public transport factors and external factors. The output of this decision is the locations chosen to perform the activities in the activity sequence.

The route choice and location choice are not simultaneously determined in the case of the unfamiliar departing passenger. These passengers do not know the station and can therefore not optimise or logically order the route and activity. In this case the route will be the result of the locations chosen to perform each activity. The route taken is influenced by the following factor groups: general personal factors, trip factors, location factors, route attributes, public transport factors and external factors. The output of the decision on route choice is the route taken when executing the activities in the activity sequence on the chosen locations.

The route chosen provides a dotted feedback line into the activity choice of the next trip. However, it is not known how the unfamiliar passenger uses this feedback or whether he uses it at all. Therefore, the status of this relationship is unknown.

2.5. CONCLUSIONS

This chapter addressed the literature review on the choice behaviour of departing passengers in the station building. First, the choices that are relevant in the station building were introduced. These are activity choice, activity hierarchy, activity sequence, activity location choice and route choice.

Second, the factors influencing the choices made in the train station were determined from literature or based on observations taken in the train station. One factor, the familiarity of a person, was found be of significant influence on the choice behaviour of pedestrains. This means that a familiar person makes choices in a different way than an unfamiliar person. This was used in this research in the creation of the theoretical framework. The factors identified were divided over four main groups: personal, system, public transport and external. The personal factors were divided into general personal, trip factors and learning processes. The system factors were divided into location factors and route attributes.

The literature search provides interesting research opportunities for several of the factors identified. This was due to no (quantified) research available. These are the time spent in the station building, time of day or week, orientation, visibility, travel time, distance, time table and train operations.

Third, the relationships between the choices made in the train stations were identified from literature. Several frameworks were identified that could help in the creation of the theoretical framework. These frameworks also came from pedestrian behaviour studies. These frameworks identified two levels that are important in pedestrian choice behaviour: before trip and during trip. A pedestrian makes plans before the trip (usually seen as the planning of activities) and executes a certain plan (including adaptions) during the trip. Here he determines the location, route and schedule of the activities.

Finally, the theoretical framework was created. Due to the significant difference expected in the choice behaviour of familiar and unfamiliar pedestrians, two frameworks are introduced (familiar versus unfamiliar passengers). These frameworks include decisions made on two levels: before trip and during trip. The familiar passenger knows the station and can therefore plan further ahead than the unfamiliar passenger. Therefore the familiar passenger plans all aspects of his trip on beforehand. He then adapts the sequence, route and locations of activities when needed (could be due to e.g. pedestrian operations, later arrival time, impulse behaviour or a delayed train). The unfamiliar passenger is only able to plan the activities, assign importance to these activities and plan a basic activity sequnce before the trip. During the trip he will determine his actual activity sequence, route and locations.

The relationships identified in the theoretical framework form the basis for the quantitative study that will be testing the extent to which these factors influence the choices made in the station. However, the theoretical framework covers a large amount of relations. These can, due to time constraints, not all be tested in this research. Therefore a selection of the relationships concerning route choice and activity location choice must be made. This is done in Chapter 3.

3. SELECTION OF CHOICE BEHAVIOUR FOR THIS STUDY

Chapter 2 identifies the theoretical framework. Not all of the relationships in the framework are researched in the state-of-the-art or can be captured with the data collection method SMART station. Therefore it is necessary to make a selection in the relationships that are researched in this study. Route choice and activity location choice behaviour have a central role in the objective of this research. Therefore, only the choice behaviour related to route and activity location choice is taken into account for this selection. Since not all relationships can be captured by means of data collected with the SMART station method, it is necessary to assess the suitability of the method for this research. The following question is addressed in this chapter:

'(3) To what extent is SMART station a suitable data collection method for collecting quantitative data on route and activity location choice behaviour of departing pedestrians?'

First, SMART station (the data collection method) is introduced in section 3.1 and the theoretical assessment on the suitability of the method is discussed. Section 3.2 addresses the selection criteria on which the relationships are selected. Here the practical suitability of the method is discussed. Finally, conclusions are derived in Section 3.3.

3.1. DATA COLLECTION METHOD - SMART STATION

As mentioned before, the data collection method used in this research is SMART station. It is necessary to assess whether it is a suitable method to capture the route and activity location choice behaviour of departing pedestrians in train stations. The theoretical assessment is addressed in this section. This is done by providing an explanation on how the method works (Section 3.1.1), comparing the method to other data collection methods (Section 3.1.2) and listing its advantages and limitations (Section 3.1.3).

The concept of the SMART station method consists of tracking and counting pedestrians (see Figure 12). This is done by means of scanners: Bluetooth and Wi-Fi scanners for tracking and infrared scanners for counting. This means that the scanners used for tracking only capture people who carry a device with Bluetooth or Wi-Fi enabled. The infrared scanners, on the other hand count all the people that pass by the scanner.

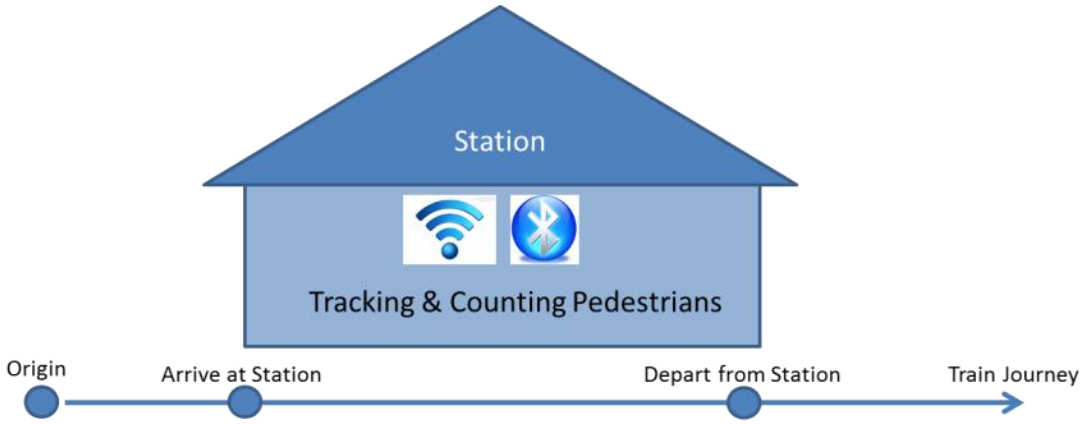

Figure 12: SMART station concept

The Bluetooth and Wi-Fi scanners are placed inside the station building and track people from the moment they arrive at the station building until they leave. The amount of scanners and their distribution over the station building determines the preciseness and extent to which people can be tracked. Because not all people have devices that emits Bluetooth or Wi-Fi signals, only part of the population that visits the station can be captured. The infrared scanners count all the people that visit the station and can therefore be used to calibrate the tracking scanners.

The data collected by SMART station can be used to identify movements made by pedestrians (see Figure 13). As can be seen in Figure 13 it is not possible to identify the exact movements of a person in the station building. The choices that can be identified by means of SMART station data are the routes passengers completed (via all activities), the activities they performed, the sequence of the activities and the location of the activities. Again, the extent to which the movements can be captured depends on the amount of scanners and their distribution in the station.

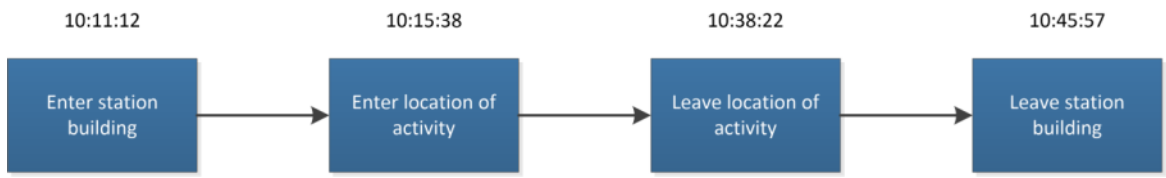

Figure 13: Example of representation of choices made by SMART station

Several stations in the Netherlands are equipped with the SMART station method: Utrecht Central station, Schiphol and Leiden. Utrecht Central station has the largest coverage in the station building (amount of scanners and distribution). This means that this station provides the most precise data of all the stations that are equipped. Therefore Utrecht Central station is used for the case study (see section 4.1 for more information on this station).

3.1.1. DATA COLLECTED WITH SMART STATION

The data collected by means of SMART station are individual data, meaning that each person tracked by the method represents one case. The data consist of several aspects: start time and date, end time and date, user number, punch card and activity sequence. The start time and date describe the moment someone enters the station building. This can be via one of the entrances or via the stairs or escalators providing access to the station building from the platforms.

The end time and date describe when someone is not seen by any scanner for a time period of 1 hour. The limit is set to one hour, because a traveller going out of the station building for a short time period should have only one complete route. The route is then terminated at the scanner where the person was last seen.

The user number is unique for each device (smartphone, tablet, laptop etc.). By means of this number a pedestrian is recognised during his or her route in the station. Each scanner receives signals from the device of that pedestrian. A pedestrian is provided with a new user number every day he enters the station. This way the privacy of the pedestrians is respected (van den Heuvel et al., 2013).

The punch card is introduced to capture the movements a pedestrian makes in the station building. A punch represents an activity performed by a pedestrian. The punch card differs per station, as every station has a different lay-out and scanners at different locations. For Utrecht Central station a total of 69 possible punches are identified. These are used to capture all the people present in the station with a device enabled with Wi-Fi or Bluetooth (departing, arriving, transferring and non-passengers). The punches are not equal to the locations of the scanners. Some scanners are used individually and others are combined into activities. An example for the former is the scanner in the AH to Go. An example of the latter is leaving via platform 11/12, this platform can be accessed via three stairs and includes three scanners. An overview of the possible punches at Utrecht Central station is presented in Table 6.

Table 6: Punch card for Utrecht Central station

Punch	Number	Punch	Number
Enter via Hoog Catharijne	0	Transferring passenger	35
Enter via Jaarbeurs	1	Non-passengers	36
Enter via Bus platform (East)	2	Stairs 18/19 South Down	37
Enter via Platform 11/12	3	Stairs 18/19 Mid Down	38
Enter via Platform 14/15	4	Stairs 18/19 North Down	39
Enter via Platform 18/19	5	Stairs 14/15 South Down	40
Enter via Platform 5/7	6	Stairs 14/15 Mid Down	41
Enter via Platform 8/9	7	Stairs 14/15 North Down	42
Enter via Platform 1-4	8	Stairs 11/12 South Down	43
Leave via Hoog Catharijne	9	Stairs 11/12 Mid Down	44
Leave via Jaarbeurs	10	Stairs 11/12 North Down	45
Leave via Bus platform	11	Stairs 8/9 South Down	46
Leave via Platform 11/12	12	Stairs 8/9 Mid Down	47
Leave via Platform 14/15	13	Stairs 8/9 North Down	48
Leave via Platform 18/19	14	Stairs 5/7 Mid Down	49
Leave via Platform 5/7	15	Stairs 5/7North Down	50
Leave via Platform 8/9	16	Stairs 18/19 South Up	51
Leave via Platform 1-4	17	Stairs 18/19 Mid Up	52
Shop Rituals	18	Stairs 18/19 North Up	53
Shop Smullers	19	Stairs 14/15 South Up	54
Shop Starbucks (BK)	20	Stairs 14/15 Mid Up	55
Shop Burger King	21	Stairs 14/15 North Up	56
Shop Julias	22	Stairs 11/12 South Up	57
Shop Tickets and service	23	Stairs 11/12 Mid Up	58
Shop Starbucks	24	Stairs 11/12 North Up	59
Shop Broodzaak	25	Stairs 8/9 South Up	60
Shop AH to go (8/9)	26	Stairs 8/9 Mid Up	61
Shop AH to go (Etos)	27	Stairs 8/9 North Up	62
Shop Paperchase	28	Stairs 5/7 Mid Up	63
Shop Kiosk 5/7	29	Stairs 5/7 North Up	64
Shop Kiosk 8/9 North	30	Depart via train	65
Shop Kiosk 11/12 North	31	Arrive via train	66
Shop Kiosk 8/9 South	32	Bluetooth user	67
Departing passenger	33	Wi-Fi user	68
Arriving passenger	34		

The types of activities that can be identified are the following: enter or leave the station via entrance, enter or leave the station via platform, visit a shop (15 shops equipped with the SMART station method), the type of pedestrian, use the stairs up or down, arrive or depart via train, use Bluetooth or Wi-Fi.

The punch card does not provide a sequence in the activities. This is due to the binary nature of the punch card. It only shows whether a pedestrian was seen at that activity or not. If more than one activity is performed the sequence cannot be derived from the punch card. The activity sequence is therefore introduced.

3.1.2. COMPARISON OF DATA COLLECTION METHODS

When collecting data the first decision is whether these are collected in real life or in a laboratory setting. (Bovy & Stern, 1990). This is translated to no control on other possible influences in real life versus a controlled setting where one can test the factors of interest without intervention of other influences in a laboratory setting. Related to the setting of the data collection is the type of data that are collected: stated preference or revealed preference (Bovy & Stern, 1990). Stated preference data is mainly used to capture information and data about future or possible situations, which (often) cannot be captured in real life. This can therefore be defined as a laboratory setting. Revealed preference data are used to capture real life information: what people actually did, chose or experienced. A drawback of this method is that one cannot easily change the conditions to see how the change affects the choice behaviour. This could imply limitations to the research. However, in this study the focus is on knowing how people behave in the train station in a 'normal' situation, meaning that the situation does not need change. The SMART station method provides revealed preference data, thus the decisions made by people.

In literature several data collection methods are used that can be compared to the SMART station method:

- ***Direct observation*** → Helbing et al. (2002) collected empirical data on pedestrian crowds for over four decades. One of the methods they used was direct observation. This can be done in two ways. The first is *local observation* where the researcher observes the people passing by (e.g. Helbing et al., 2002 and Daamen & Hoogendoorn, 2003). This method can for example be used to count flows. The SMART station method can also be used to count flows. The second is *stalking* people and thus following them on their route (Hill, 1982). Applying this method would imply that one can trace the exact route of a person. This means that this method could be more precise compared to the SMART station method. Direct observation is exposed to human errors, which can result in miscounting. It is a very time consuming method and captures only a small sample of the population (especially stalking). The samples that can be collected in one period using SMART station versus direct observation differ greatly in size. The SMART station method collects much more data and requires much less effort.

- ***Survey*** → Seneviratne & Morall (1985) used a survey to collect data on route choice influences. A survey can be both revealed preference and stated preference. It is revealed preference if the researcher asks people what they did, chose or experienced. A stated preference survey contains possible choices people can make that cannot be captured in real life (yet). Seneviratne & Morall (1985) asked people to indicate why they chose the route they used (revealed preference). Hill (1982) also used this method to collect data on route choice. A survey can capture much more than only the route choice of a person, for example also the socio-demographic information of each person. The level of detail that can be asked by means of a survey is much higher than is possible with SMART station. However, errors arise when asking people about the route or activities they performed. People could provide the answer that is socially accepted or they could not remember the exact route or activities. This is a drawback of this method. When actually observing people (revealed preference) this error does not occur.

- ***Photos*** → Helbing et al. (2002) also made use of photos of multiple locations. Photographs of an area can be used to identify the density in that area at several moments in time. This is similar to one of the possibilities of SMART station. In the entire station building one can count how many people are present at one point in time (using the infrared scanners). A photograph visually shows the actual density, but needs a lot of time processing the data collected.

- ***Video cameras*** → Lee et al. (2001) collected data on the route choice of a pedestrian by means of video cameras. They observed movements of pedestrians through a station to deduce which route pedestrians took through the station. This same method was applied by Antonini et al. (2006). This method captures (nearly) everyone that passes the camera(s) and provides a visual image of the route taken. The camera footage of multiple cameras needs to be combined and the same person needs to be identified during the entire route. This configuration is very time consuming. Next to that, mistakes can be made when combining the cameras. Human error increases with the use of multiple cameras. Also, cameras on different angles capture different images. The method needs a translation to data that can be easily interpreted. Summarised, video cameras are able to capture the route of a pedestrian more precisely than SMART station (due to visual footage). However, the configuration and translation of the camera footage into usable data is very time consuming and prone to human errors. These drawbacks are not found at the SMART station method.

- ***Video time-lapse movies*** → Helbing et al. (2002) also used the method video time-lapse movies. These movies can be used in a similar way as photographs: for determining the density at one location. Video time-lapse videos are also used by Cheung & Lam (1998). The advantage of video-time lapse movie compared to a normal movie is that the image is captured at a fixed frequency (e.g. every 15 seconds). This results in a lower capacity needed to store the data.

- ***Bluetooth tracking*** → Versichele et al. (2012) use a more recently introduced method to collect data: Bluetooth tracking. This method follows devices, like mobile phones, that have their Bluetooth function enabled. This way they collected data on the spatiotemporal movements of people during a mass event (the Ghent festivities). Not everyone has their Bluetooth enabled, the capture rate was approximately 11%. SMART station also uses Bluetooth tracking, but combines this with more data collection methods (Wi-Fi and infrared).

- ***Smartphone data*** → Research by the central service of statistics (CBS) in the Netherlands on the use of smartphones as a data source for research related to mobility provided interesting insights related to the possibilities of smartphones (Roos & Arends-Tóth, 2013). A smartphone can produce around 15 different sources of data. The ones that are seen in research are GPS, Wi-Fi and Bluetooth.

Several data collection methods have been identified and compared with the SMART station method. The main advantage of SMART station is the large amounts of data it collects with minimal effort. Other methods are not able to generate the same amounts of data on choice behaviour. However, these methods sometimes produce more accurate details (e.g. video cameras that show the route) or extra data (e.g. survey that collects socio-demographic data) that SMART station cannot collect.

3.1.3. LIMITATIONS OF SMART STATION

As mentioned before, the main advantage of SMART station is its ability to generate data on many pedestrians with minimal effort required from the researcher. However, some limitations are present in the data collected with the SMART station method.

When a pedestrian has a device with Bluetooth or Wi-Fi enabled, he emits a signal that is seen by each scanner with a certain strength: RSSI (Received Signal Strength Indication). This RSSI ranges from -100 dBm to 0dBm, with zero being a perfect signal and -100 being the worst signal. Using the distances between the device of the pedestrian and the scanners of SMART station, the location of the pedestrian can be determined. Each scanner receives a signal from the device with a certain strength that decreases with increasing distance. This way the location of a pedestrian can be determined (highest strength is closest to actual location). However, when the scanner closest to that pedestrian is not working the second closest scanner receives the highest signal strength. In that case an error is introduced to the location specification of the pedestrian. After looking into the log of all the nodes at Utrecht Central station during the period 30th of August 2013 to 19th of September 2013 the conclusion is that this type of error did not occur in the data collection.

A device enabled with Wi-Fi sends (in the worst case) a signal only once every 45 seconds. This frequency depends on the type and brand of the device (in case of a smartphone: I-phone is known to be the worst case with 45 seconds). This means that a pedestrian, if he is fast, could perform more activities than recorded. Bluetooth on the other hand, sends a signal every nearly every second. This means that Bluetooth is much more precise than Wi-Fi. However, a larger part of the population uses Wi-Fi compared to Bluetooth: 91.5% versus 8.5% in the dataset of Utrecht Central station. The large share of Wi-Fi use combined with the frequency of the signal influence the reliability of the data collected. In order to reduce the reliability issue, the pedestrians that have a very low time spent in the station building are excluded from the dataset. The limit is set to 45 seconds, this covers the worst case scenario. This is only 0.6% of the total dataset, therefore the impact on the dataset is small. It should be kept in mind that Wi-Fi is less accurate than Bluetooth and can introduce an inaccuracy of on average 10-20 seconds with a maximum of 45 seconds and a minimum of 0 seconds.

How much time does a pedestrian need to spend inside a shop in order to be a shopper? Given the transaction time and time searching for the correct product within a shop this limit is set to 30 seconds. The minimum transaction time in the AH to Go is around 22 seconds. A shopper then also needs to search for his product, get it and sometimes stand in line. This could mean that if someone is very fast, they are not counted as shoppers and a misinterpretation could occur. But because the non-shoppers need to be separated from the shoppers, this time limit seems plausible. The probability that someone is faster than 30 seconds and is a shopper is very low.

Summarised, the SMART station method could have several limitations related to the reliability of the data. However, in the case of the data collected at Utrecht Central station most of these issues were small. Misinterpretation of the location of a pedestrian could not occur because all the scanners were working during the data collection period. Shoppers can be missed when they are too fast. However, the limit for a shopper seemed plausible at 30 seconds because of the need to separate them from non-shoppers. The accuracy of the data collected based on the frequency of the signal emission is high in the case of Bluetooth and lower in case of Wi-Fi. The latter depends on the type and brand of the device. The accuracy of the data collected however needs to be taken into account in the remainder of this research.

3.2. SELECTION CRITERIA

In this section the selection criteria used for the quantitative research is addressed, see Figure 14. The start of the selection process were all relationships present regarding activity location choice and route choice. This means every factor that could influence either of these choices. In Chapter 2 the list of all possible relationships was reduced to the factors (and relations) observed in literature and practice that are relevant for train stations. However, this list is still quite extensive. Therefore, the relationships are also tested on the knowledge available and on the possibility to capture the relationships using SMART station data. If no knowledge is available, opportunities arise for this research. Since the method used to collect data is the SMART station tool, the relationships that are selected should be captured by these data. This provides the practical assessment of the SMART station method as a suitable data collection method. In other words, the method can be used to capture the most interesting relationships identified.

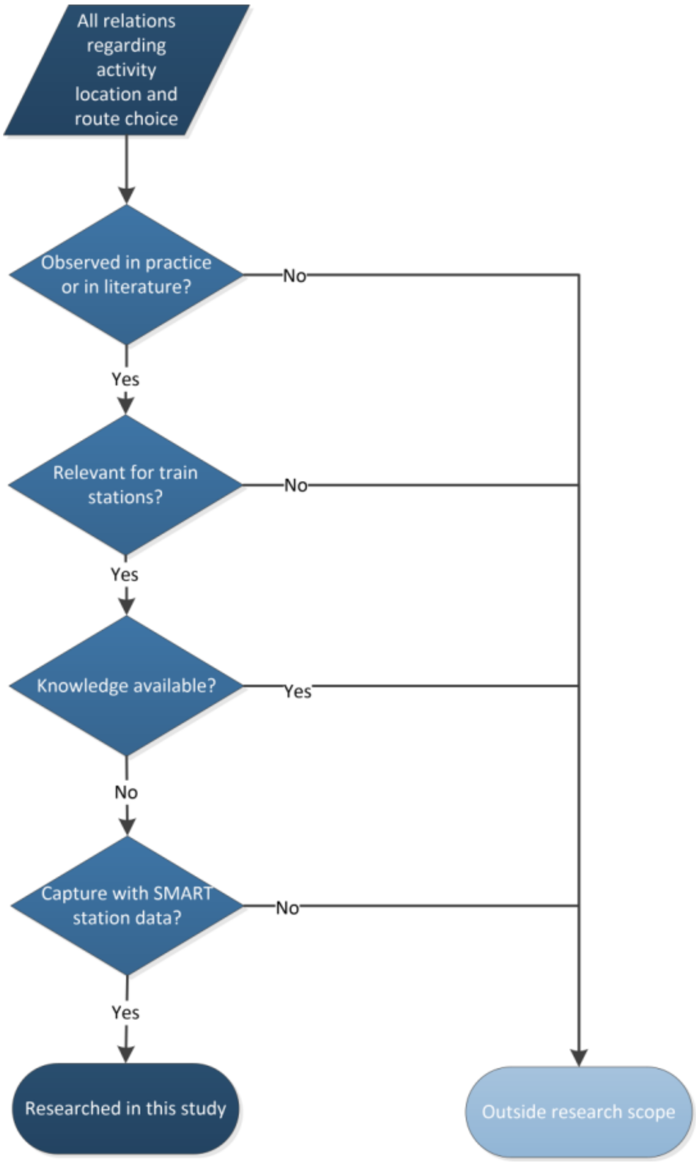

Figure 14: Flowchart with selection criteria

3.2.1. AVAILABILITY OF KNOWLEDGE

The first criterion on which the relationships are tested, after the creation of the theoretical framework for departing pedestrians in train stations, is the availability of knowledge. There are two sub-criteria in determining if knowledge is available: whether it is researched in the state-of-the-art and whether it is quantified in the state-of-the-art. This is summarised in Table 7.

This criterion is applied as a knock-out criterion. If the relationship is already researched in the state-of-the-art and also quantified, then a lot of knowledge is already available. Therefore the added value of researching that relationship in this study would be low. This relationship is then outside the scope of this research. If the relationship is researched in the state-of-the-art but was not quantified, interesting opportunities arise for this study. When the relationship is not quantified this could be due to limitations of the availability of data or the budget available. If the relationship is not researched in the state-of-the-art, it could be observed in practice and therefore offer interesting research opportunities. An extensive overview of the outcome of the test with this criterion per relationship can be found in Appendix B.

Table 7: Availability of knowledge

		Quantified in the state-of-the-art	
		Yes	*No*
Researched in the state-of-the-art	*Yes*	Yes	No
	No	x	No

3.2.2. USABILITY OF SMART STATION DATA

The relationships that passed the test on availability of knowledge are tested with another criterion. This is whether they can be captured using SMART station data. The relationship can be captured by means of SMART station if the factor, choice and the relationship itself can be captured. If this is not the case then the relationship cannot be captured using SMART station data. The relationships that remain after this selection are shown in Table 8. An overview of all the relationships tested can be found in Appendix C.

Table 8: Relationships that can be captured using SMART station data

Relationships that can be captured with SMART station	
Orientation	Route choice
Orientation	Activity location choice
Time spent in the station	Route choice
Time of day/week	Route choice
Time spent in the station	Activity location choice
Time of day or week	Activity location choice
Travel time	Route choice
Distance	Route choice
Travel time	Activity location choice
Distance	Activity location choice
Visibility	Route choice
Time table	Route choice
Train operations	Route choice
Time table	Activity location choice
Train operations	Activity location choice

The orientation reflects the choice for left or right. In the Netherlands many things are performed on the right side: e.g. walking or driving. This orientation can be captured by looking at the map of the station building and reflecting the orientation of the route or location of the activity on the SMART station data of the individuals.

The time spent is registered by the SMART station method. The route chosen is logged using the punch card and activity sequence of the method. This however happens in a binary nature. The exact movements are not known, but the route can be determined via all activities performed. Also, the location of the activity and the time of day or week are also registered by the method.

The travel time can be derived from SMART station data. The travel time can be defined as the walking time of a person, whereas the time spent also consists of the time spent inside a shop. The distance covered cannot directly be derived with the SMART station method. But the combination of measurements from a map of the station building and SMART station can be used to capture the relationship between distance and route choice and activity location choice.

The visibility relates to the visibility of the route or parts of the route. This is not measured directly by means of SMART station, but it can be determined from a map of the station building. Therefore the relationship between route choice and visibility can be determined.

The time table consists of e.g. frequency, platform and service type. This information is not measured with SMART station. However, it can be gathered at the Netherlands Railways (NS) because all the train movements are registered. The same is true for the train operations. The data can then be matched or combined with the SMART station data to help capture the relationships between time table and train operations with route choice, activity location choice and activity sequence.

3.2.3. SUMMARY ON THE SELECTION CRITERIA

The assessment of the relationships identified in the theoretical framework with the criteria availability of knowledge and usability of SMART station data, can be summarised with an overview provided in the shape of a matrix (see Table 9). The matrix shows a score for each relationship the score for both criteria. The relationships that provide interesting research opportunities for this study are provided in the top left corner of the matrix. These relationships can be captured by means of SMART station and no knowledge is available yet.

Table 9: Overview of assessment on the criteria

		Capture using SMART station data?	
		YES	NO
Knowledge available?	NO	Time spent - Route choice Time of day/week - Route choice Time spent - Activity location choice Time of day or week - Activity location choice Travel time - Route choice Travel time – Activity location choice Distance - Route choice Distance - Activity location choice Time table - Route choice Train operations - Route choice Time table - Activity location choice Train operations - Activity location choice Visibility – Route choice Orientation – Route choice Orientation – Activity location choice	Emotional state - Route choice Personal uncertainty - Route choice Personal uncertainty - Activity location choice Group composition - Route choice Trip purpose - Route choice Impulse behaviour - Route choice Learning processes - Route choice Learning processes - Activity location choice Location dimensions - Activity location choice Information - Activity location choice Amount of shops - Activity location choice Route attributes - Route choice Crowdedness - Activity location choice
	YES	Familiarity - Route choice Familiarity - Activity location choice Time of day or week - Activity location choice Waiting area - Activity location choice Weather - Activity location choice Weather - Route choice	Age - Route choice Gender - Route choice Location dimensions - Route choice Route constraints - Route choice Information - Route choice System uncertainty - Route choice System uncertainty - Activity location choice

In the ideal situation the SMART station data collection method would be able to cover all the relationships identified in the theoretical framework, of which no knowledge is available yet. This however is not the case because the information needed for testing some of the relationships is not measurable with the SMART station method.

The fact that not all interesting relationships can be tested using data collected with SMART station means that the method is not completely suitable. The relationships that cannot be tested are mainly related to socio-demographic and personal aspects. Therefore, with a combination of the SMART station data and personal data it would be possible to test all relationships identified. However, because of the privacy that needs to be respected when collecting data with SMART station, it is not possible to link the data collected to individuals. When using other data collection methods that can collect personal data, the amount of pedestrians involved in the research will be severely lower than with the SMART station method. In addition, this would mean that other relationships of interest cannot be captured. The relationships that can be measured and tested do also provide interesting information for both science and practice. Therefore, the use of this method is plausible.

3.3. CONCLUSIONS

The selection of the choice behaviour that will be researched in this study is based on the relationships identified in the theoretical frameworks (Chapter 2). The selection is based on two criterion: availability of knowledge and the usability of the data collected by means of SMART station. SMART station is the method used to collect data on the choice behaviour of the departing passengers. The method uses the Wi-Fi and Bluetooth technology available on smart phones to track pedestrians through the station building. In addition, the method uses infrared scanners to count people in the station. The infrared scanners are also used to calibrate the data collected on Wi-Fi and Bluetooth because not everyone in the station has a device with either of these technologies enabled.

The method is able to collect data on many passengers with relatively minimal effort required from the researcher. When comparing SMART station to other data collection methods, several conclusions were derived. Other methods are not able to collect the amount of data that SMART station can. Some methods are more detailed on the exact movements of pedestrians (video footage or stalking). Other methods also collect personal data on the pedestrians (survey), which is not possible with SMART station.

The suitability of the SMART station data for this research was assessed by means of a theoretical comparison and the usability of the data on the interesting relationships between to route choice and activity location choice. Theoretically the advantages outweigh the limitations, depending on the research objectives identified. If research is aimed at personal data or microscopic movements, the method is not suitable. However, this is not necessarily the case for this research. Therefore, the method seems theoretically suitable.

The practical suitability needs to be assessed after the quantitative analysis is performed. This quantitative analysis will be executed for the following relationships:

	Route choice	Activity location choice
Time spent in the station hall	X	X
Time of day or week	X	X
Distance	X	X
Travel time	X	X
Time table	X	X
Train operations	X	X
Visibility	X	
Orientation	X	X

Chapter 4 addresses the research approach of the quantitative study. The case study location, data preparation and validation and analysis method are discussed. Chapter 5 then addresses the route choice model and analysis and Chapter 6 addresses the activity location choice model and analysis.


Utrecht

4. INTRODUCTION TO THE CASE STUDY AT UTRECHT CENTRAL STATION

In the previous chapters the theory for this research was developed. The frameworks developed were assessed by testing the selected relationships on a real life case: Utrecht Central station. The daily business in the train station is the coming and going of the trains. This means that understanding this process is essential for understanding the choice behaviour of departing pedestrians. Utrecht Central station and its train processes are addressed in Section 4.1.

SMART station collects data using different scanners in the station building. These individual scans need to be combined in order to create a route. This means that a process of combining and filtering is needed before a usable dataset was found. Also as mentioned in Chapter 3, the data related to the train processes need to be combined with SMART station data. These preparations of the dataset are described in Section 4.2. The resulting dataset is tested and validated in section 4.3.

After this the method used to perform the quantitative analysis on the selected relationships is introduced in Section 4.4. As mentioned before, the method used is discrete choice analysis. The following question is addressed in this section:

'(4) Which types of discrete choice models belonging to the logit family can be used to estimate models on route and activity location choice behaviour of departing pedestrians?'

The most suitable models regarding route and activity location choice are introduced. The chapter is finalised in Section 4.5 with conclusions on the approach of the theory assessment.

4.1. UTRECHT CENTRAL STATION

Utrecht Central station is the second biggest train station in the Netherlands. Utrecht is located in the centre of the Netherlands, which makes the station a large hub for trains (see Figure 15, black circle).

Figure 15: Railway map of the Netherlands of 2013 (Treinreiziger.nl, 2013)

Attached to Utrecht Central station is 'Hoog Catharijne'. This is a large shopping mall connecting the station to the city centre of Utrecht. There are also lots of shops inside the station. Most of these shops are catered for the 'to go' segment. This means that the in-store time is as small as possible. Examples of these shops are the AH to Go, Broodzaak and Smullers. Other shops are related to the 'to stay' segment. As mentioned before, stations have changed from a place to enter the train to a place to stay and meet people. The 'to stay' shops help create this. Examples of these shops at Utrecht Central station are Starbucks and Burger King. On the other side of the station is the 'Jaarbeurs'. This is a large event complex where many events take place annually.

Utrecht Central station had the second largest passenger throughput in 2011 (NS, 2011). Every day around 230,000 people travel via this station: +/- 170,000 departures and arrivals (return trip) and +/- 60,000 transfers. The majority of the people departing Utrecht do this in off-peak hours (around 59%). For departing passengers the evening peak is higher than the morning peak (30% vs 11%). The station building itself is visited by around 140,000 travellers per day (NS Stations, 2013). The other 30,000 travellers travel via the 'Noordertunnel', which is located on the North side of the station building. This tunnel provides a route which does not pass the station building. The majority of the people (around 75%) travel for work or school related purposes (familiar passengers). The other people are mainly travelling for social-recreational purposes (unfamiliar passengers) (NS, 2011).

All the entrances and exists of the station building are equipped with the SMART station tool. However, the 'Noordertunnel' is not equipped. This means that the people travelling via this tunnel are not counted or tracked by SMART station.

4.1.1. THE LAY-OUT OF THE STATION BUILDING

In section 3.1.1 the punch card of Utrecht Central station was introduced. It was shown that 15 shops are equipped with the SMART station tool. Figure 16 visualises the location of the SMART station scanners in Utrecht Central station. The green rectangles represent scanners in the central hall and the blue rectangles represent the scanners at the entrances.

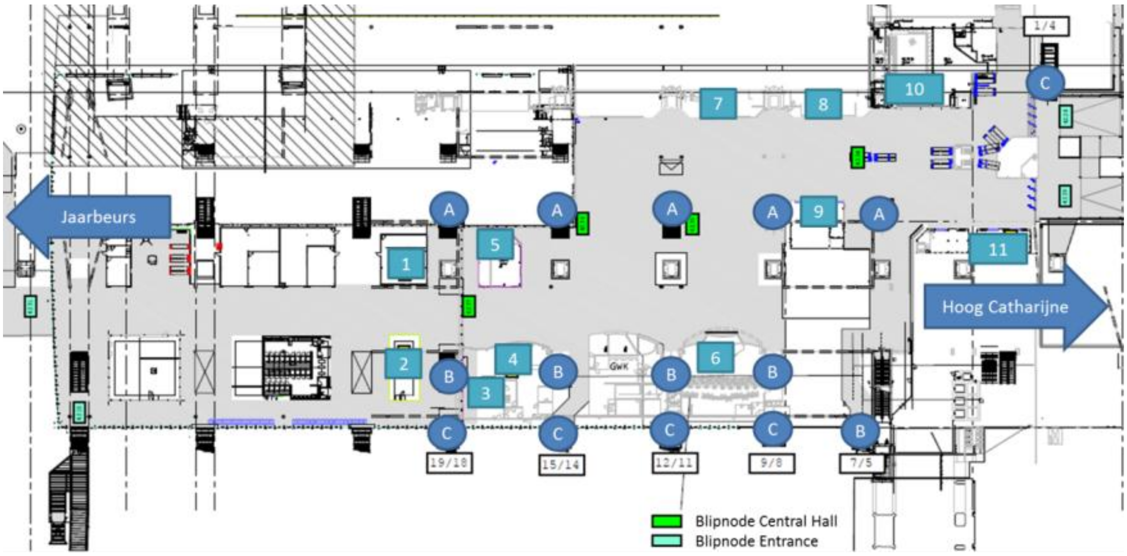

Figure 16: Map of Utrecht Central station

The following shops are equipped with SMART station (indicated by a number in Figure 16):

1. Smullers
2. Rituals
3. Starbucks (BK)
4. Burger King
5. Julia's
6. Tickets & Service
7. Starbucks (Bruna)
8. Broodzaak
9. AH to Go (8/9)
10. AH to Go (Etos)
11. Paperchase

Utrecht Central station consists of six platforms, each providing access to two or more tracks. The stairs are indicated with a C in Figure 16, whereas the escalators are indicated with A and B. Platform 1-4has a different lay-out than the others. The tracks belonging to this platform end at Utrecht Central station. These are located at the north side of the station, therefore the stairs are also located at that side of the station building. Also, instead of three access points, this platform has one. There is an extra escalator present, but this serves only arriving passengers. Platform 5/7 also has a slightly different lay-out considering the access points to the platform. Due to the access to the 'Katreinetoren', which is located next to escalator B, only two access points are present. The 'Katreinetoren' is an office of the Netherlands Railways, which is accessed from within the station building.

4.1.2. THE TRAIN PROCESSES

Departing pedestrians visit the station building with one purpose: boarding the train. Therefore, the train processes form the basis of the decision making of these pedestrians. Therefore it is important to understand the train processes for Utrecht Central station. An analysis of the processes per platform, disruption present in the period of data collection, train directions per platform, the train schedule and the train services present can be found in Appendix D. This section covers the most important findings related to these processes.

By means of the analysis of the train processes of each platform, it can be determined which platforms resemble one another. Platform 1 and 6 are exceptions to the other platforms due to their function as final and start station. Therefore they can be expected to have different behaviour of departing passengers when it comes down to for example stair usage. Next to that they serve mainly sprinter services compared to mainly intercity services on the other platforms.

Platforms 2 and 5 serve both intercity and sprinter services and have the highest number of train departures per hour. In addition, they both had a relatively high percentage of train disruptions during the three week observation period. This means that these platforms show a less constant picture than for example platform 4. Platform 5 also has the problem that at four times per hour two trains depart at exactly the same time from the two tracks. As mentioned in Chapter 3, the train process data and SMART station data need to be matched in order to test the influence of the time table and train operations on the route and activity location choice. If trains depart from the same platform at the same time this can cause problems in the matching process.

Platforms 3 and 4 serve only intercity services, which makes the behaviour of departing pedestrians on these platforms more constant over time. They have the lowest number of train departures per hour. The percentage of train disruptions on platform 4 is much lower than on platform 3. Therefore platform 4 shows the most constant picture.

Summarised, the platforms can be divided into three categories: final/start platforms, intercity platforms and mixed platforms. Each has different characteristics and serves different train services that go in different directions.

4.2. PREPARATION OF DATASET WITH ALL DEPARTING PASSENGERS

The data collected with SMART station consist of individual scans from each device. These scans need to be combined into routes. The data combining and filtering algorithm that is applied on the data collected by SMART station is addressed in section 4.2.1. After the data is combined and filtered into a usable dataset, the data needs to be combined with the train process data. As mentioned before, this is necessary to test the influence of time table and train operations on route choice and activity location choice. Each pedestrian that is tracked with SMART station needs to depart by train. This is addressed in section 4.2.2.

4.2.1. DATA SELECTION

The data collected with the SMART station tool needs to be combined and filtered in order to create a usable dataset. The algorithm for the creation of this data set is shown in Figure 17. Several steps are taken before the usable dataset is identified.

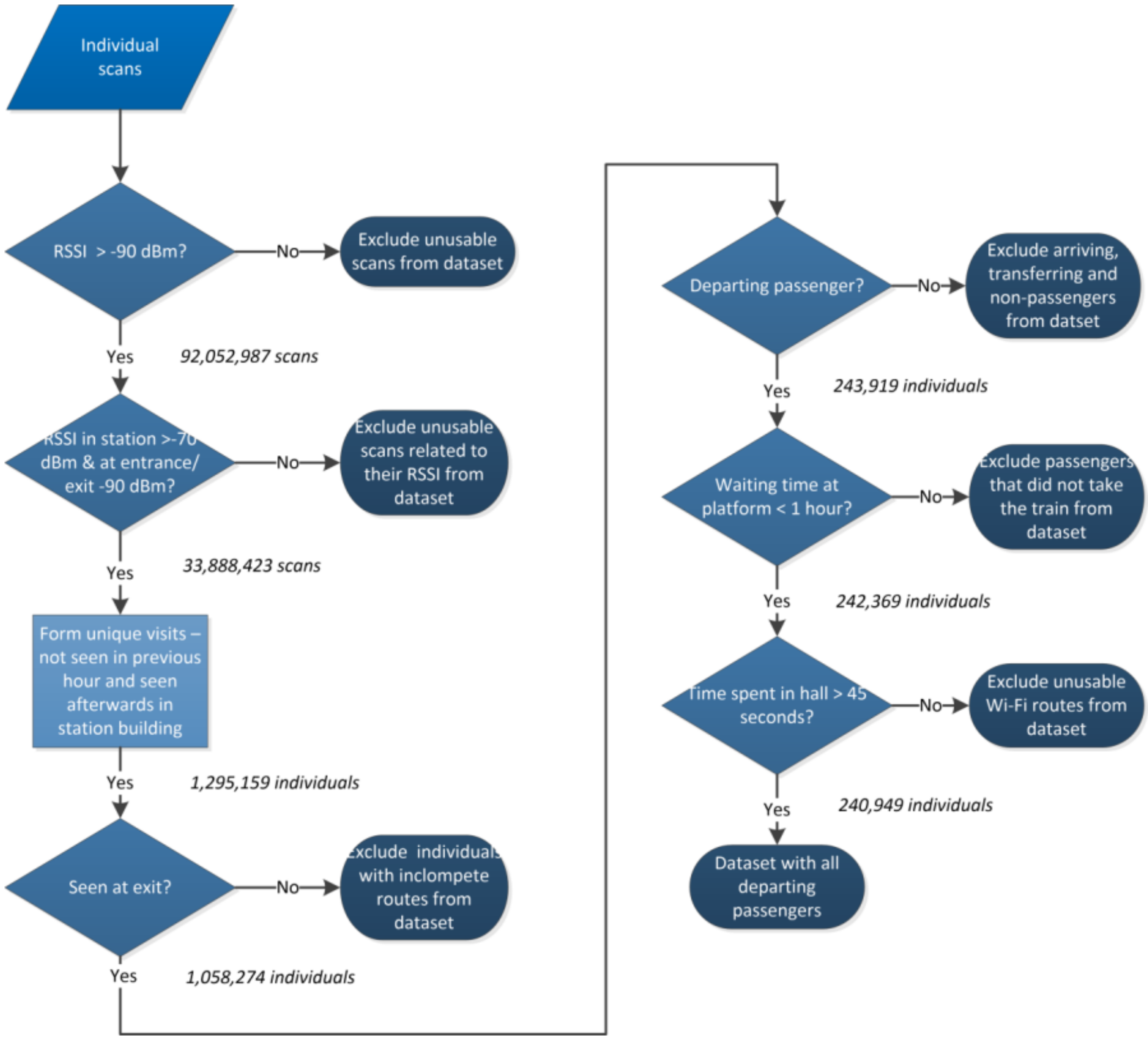

Figure 17: Data selection algorithm

To start with, all individual scans from the SMART station tool are used. As mentioned before the RSSI ranges from -100 dBm to 0 dBm, with -100 being the worst case and 0 being the best case. The first step in the data selection is therefore to remove the scans with a low RSSI (lower than -90 dBm). After this filter around 92 million scans remain. It is not known precisely how much data was collected in total, but is expected that the first filter cuts out approximately 25% of the total collected scans (Hermansen, 2014). This would mean +/- 122 million scans in total.

The second step is to set a standard to the minimum RSSI at different locations in the station building. This is done to make sure that the reliability of the scans is high. The reliability of a scan relates to the correct combining/assigning of a scan to a route of a pedestrian. When entering or exiting the station hall the demands to the RSSI are not very high, because the odds of wrongly assigning a scan to a route of a pedestrian is much lower than inside the station. Therefore the limit is -90 dBm for the entrance and exit points. These are Jaarbeurs, Hoog Catharijne and all the stairs and escalators that provide access to the platforms. Inside the station the demand of RSSI is higher (set to -70 dBm), to reduce the chance of wrongly assigning/combining the scan to the route. After this filter around 34 million scans remain.

Then the start is made in the formation of routes, thus combining all the individual scans into routes of individuals. This starts with a unique user visit. Someone is seen at an entrance point (stairs or entrance) and has not been seen by those scanners for the last hour. Also, this person is seen by one of the scanners in the station hall afterwards. After this combining process around 1.3 million individuals remain.

The route is not complete with only an entrance point and scan in the station hall. Therefore the next step determines whether a pedestrian has fulfilled his or her route according to SMART station. This is whether someone has left the station building. Around 1.1 million individuals remain in the dataset. This forms the point where all individuals have valid routes. Therefore, the next step is to remove all the individuals that were not departing. The arrivals, non-passengers and transfers are thus removed. Around 244,000 individuals remain.

Data was collected over a time period of 21 days. This means that on average +/-11,600 departing passengers were seen by SMART station. Compared to the counts that were registered by the infrared scanners at the entrances, the departing passengers present in the SMART station data are approximately 17% of the total population. This means that the number of departures per day must be multiplied by +/- six to get the total departures per day at Utrecht Central station. This is +/- 69,600 departing passengers. In order to get the total departures and arrivals in the station hall this number is multiplied by two (a passenger usually makes a return trip). This is +/- 139,200 passengers entering and leaving the station building. This number was also found by NS Stations as was already showed in section 4.1 (NS Stations, 2013). Therefore, the SMART station method shows a plausible number of departing passengers over a period of three weeks.

The next step is to remove those who were on the platform too long (this can be determined after the train assignment, see section 4.2.2). These people probably did not take a train, but left via the 'Noordertunnel'. During the night a train departs every hour. The people that just missed the train and had to wait for nearly an hour should not be left out. Therefore the limit for waiting on the platform is set to one hour. If the passenger remains on the platform for a longer time period according to the data, he probably did not take the train. After this selection around 242,000 individuals remain.

The last selection is to make sure that no invalid or unusable Wi-Fi results are included in the dataset (see section 3.1.3). Therefore all individuals that spent less than 45 seconds in the station building are removed from the dataset. This leaves 240,949 departing passengers in the dataset.

4.2.2. MATCHING OF TRAIN PROCESS AND SMART STATION DATA

In order to estimate the influence of time table and train operations on route and activity location choice, the datasets of train processes and SMART station need to be matched. The data collected on train processes consist of a date, planned departure time, actual departure time, train number, service type and track. These data need to be matched to the data from SMART station. Figure 18 shows the data matching algorithm created to match the departure of an individual from the station building to a departing train.

To start the data matching algorithm, data is needed on all departing passengers at Utrecht Central station. Then a match is made between the departure platform of an individual and a train. At this point the list of possible trains is brought down to all trains that depart from that platform.

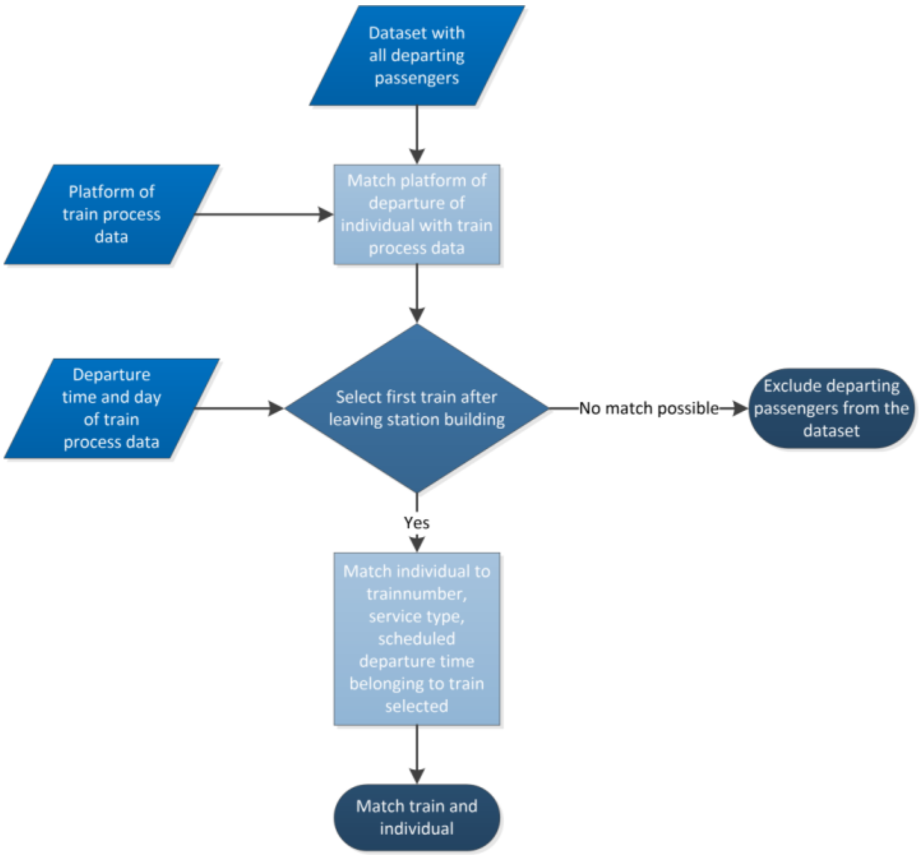

Figure 18: Data matching algorithm

Next, data is used on the time and day of the departure by the individual from the station building and the departure time and day of the trains. The departure time and day are matched in such a way that the first train leaving the station after the individual leaves the station building is matched to the individual. When this was not possible (departures on the 19th of September 2013 after the last train has left the station building) these were excluded from the dataset. In the cases where a train could be matched to the individual, the train number, service time and scheduled departure time need to be added. An example is provided to show the working of the algorithm.

A pedestrian is present in the station and departs from platform 5/7. In the train process data this platform is selected. The pedestrian leaves the station on the 5th of September 2013 at 10:55:53 AM. The first train departure on that platform after this end time is 10:58:16. This train is then matched to this pedestrian. The train number is 3132 (train series heading to Schiphol). The service type is Intercity and the scheduled departure time is 10:58:00.

DISCUSSION ON THE DATA MATCHING ALGORITHM

People do not necessarily take the first train that arrives on their platform. It is possible that a person skips the first (or even second) train because it is not heading in the right direction. One example where this is possible is the departure of the ICE train. This train has a low frequency and people wait for this specific train because they need to go to Germany. From experience and observations it is known that people wait longer on the platform to catch this train. They want to be sure they catch it. Therefore they might let one or more trains pass. In the train assignment algorithm the person is assigned to the first departing train, but not to the ICE. This same issue arises when two trains depart at the same time (platform 14/15) or when a train is delayed. This means that not all travellers have the correct train assigned to them and therefore an error can be introduced to the dataset. Due to lack of data on their final destination, the size of this error cannot be determined. Another method of train assignment could be to present a probability that a person arriving at that moment on the platform takes the next train versus the one after that. However, this would also produce an error since people can also be wrongly assigned. At the Netherlands Railways no research is available yet about the distribution of passengers over each train. Therefore, this method would not improve the train assignment. It was decided to stick to assigning the first train to the individual and acknowledge the fact that this might introduce some wrongly assigned trains.

Some pedestrians use the station building as a through fare, for example, a quick way from the city centre to the 'Noordertunnel' at Utrecht Central station. SMART station thinks these people are departing passengers, because they enter the station via an entrance and leave via a platform. However, they continue their journey through the Noordertunnel. There is no data available within NS that shows how large this number of travellers is. The consequence for the algorithm is that it assigns some people that were actually not departing at all. Therefore a slight overestimation of the number of departing passengers is present. This is especially true during the night, when no trains depart from most platforms but people seem to depart. It is known that platforms 18/19,1-4 and 5/7 are most frequently used for these movements. This will nearly never occur on the other platforms, because the route is not logical. Therefore, when analysing the other platforms, this type of error is very small.

If a pedestrian is a moment too late for the departing train, he is assigned to the next train. There is however a possibility that the last signal is received from inside the train and that the pedestrian was actually in it. In that case the next train has an overestimation of passengers and the actual train has an underestimation. This size of this error is not known, however the impact of this error is expected to be relatively small.

Some trains did not have a track assigned to them. The information on these trains was not recorded properly. Because they do not have a track assigned to them it is not possible to match them to the departing pedestrians. Therefore some trains are excluded from the algorithm. This could create larger waiting times on the platform than actually arose. In total 138 trains in three weeks did not have a track assigned out of 18,311 trains. This is 0.75%. This means that the error introduced is very small.

Finally, during some of the days no trains left on certain platforms due to engineering works. This happened on the 1st of September 2013 on platform 14/15, on the 7th and 8th of September 2013 on platform 18/19 and on the 14th of September 2013 during 01:00h and 22:00h on platform 8/9. The people departing from these platforms during these days or times were excluded from the dataset, because they were obviously not departing by train. This was 1 pedestrian on the 8th of September 2013 and 10 on the 14th of September 2013. Since they did not depart by train they were heading towards the 'Noordertunnel'. This means that the error implied by pedestrians departing via this tunnel is relatively small.

4.3. VALIDATION OF DATASET WITH ALL DEPARTING PASSENGERS

In the previous section an explanation was given on how the dataset containing all departing passengers was created. This section addresses the validity of this dataset. This is done in two parts. First, by comparing the data to other data sources available at NS Stations. The data sources used are the 'Keten Informatie Systeem (KIS)' (NS, 2011) and counts with infrared scanners at the entrances. The KIS dataset provides information on the distribution of departures over the day. The counts with the infrared scanners can be used to validate whether a representative sample from the entire population is present in the dataset. The second part in testing the validity of the dataset consists of testing differences and comparisons between Wi-Fi and Bluetooth measures, because it is expected that different types of people make use of different technologies. By testing and comparing the data collected by both technologies, it was determined whether they are indeed significantly different or not.

4.3.1. TOTAL DATASET

Several aspects of the dataset are compared to KIS and the counts of infrared scanners. These are the amount of departures per day, the amount of departures during the day, travelling during peak hours or off-peak hours, time spent in the station hall, shop visits and departures per platform.

DEPARTURES PER DAY

The period of the data collection was from the 30th of August until the 19th of September 2013. The total amount of departures over that period was 240,949 people. That capture rate for all pedestrians (also including non-passengers, arriving passengers and transferring passengers) in Utrecht Central station is estimated to be 26% (Hermansen, 2013). The entrances at Hoog Catharijne and Jaarbeurs are equipped with infrared scanners. These also capture the people that do not carry a device with Bluetooth and Wi-Fi enabled. This way it can be tested if a bias was created in the number of departures per day or not. Figure 19 shows the comparison in percentages of the departing passengers each day. An absolute comparison was not possible, since the Wi-Fi and Bluetooth data are only a subset of the total departures. When looking at the Wi-Fi and Bluetooth data as a percentage of the counts it shows that there is only a small difference between them on each day. The largest difference is found on the 19th of September 2013, because in the afternoon the infrared scanners were turned off. A Kolmogorov-Smirnov test was executed to test whether the distributions of the pedestrians over the days is significantly different. On a 95% confidence interval the test provided a significance of 0.841. This means that the distributions are significantly equal. This also means that the data collected by SMART station provides a representative sample, looking at the distribution over the days.

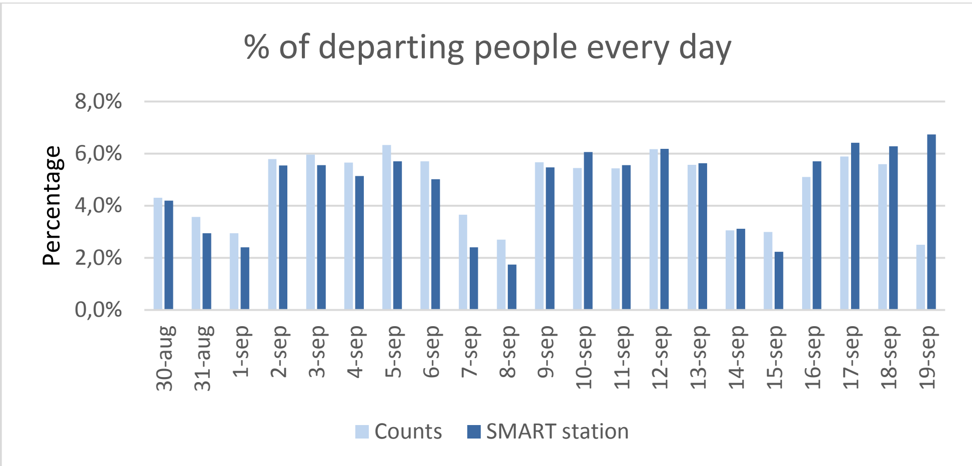

Figure 19: % of departing passengers every day

DEPARTURES DURING THE DAY

The variation during the day of Wi-Fi and Bluetooth data needs to be validated, by comparing it to the infrared counts at the entrances. Figure 20 shows the spread in percentage of departing passengers over the day. This is an aggregation of the three weeks of data. It is visible that a slight overestimation in the Wi-Fi and Bluetooth data is present during the peak hours. The people heading for work or school (mainly travelling in peak hours) obviously have a higher share in the enabling of Wi-Fi and Bluetooth than the people traveling in the off-peak hours. The distribution of both the counts is compared with the distribution of SMART station data by means of a Kolmogorov-Smirnov test. This tests the equality of the distribution. The significance value is 0.893 on a 95% confidence interval. This means that both distributions are significantly equal. The distribution of pedestrians over the day collected by SMART station is thus representative of the entire population.

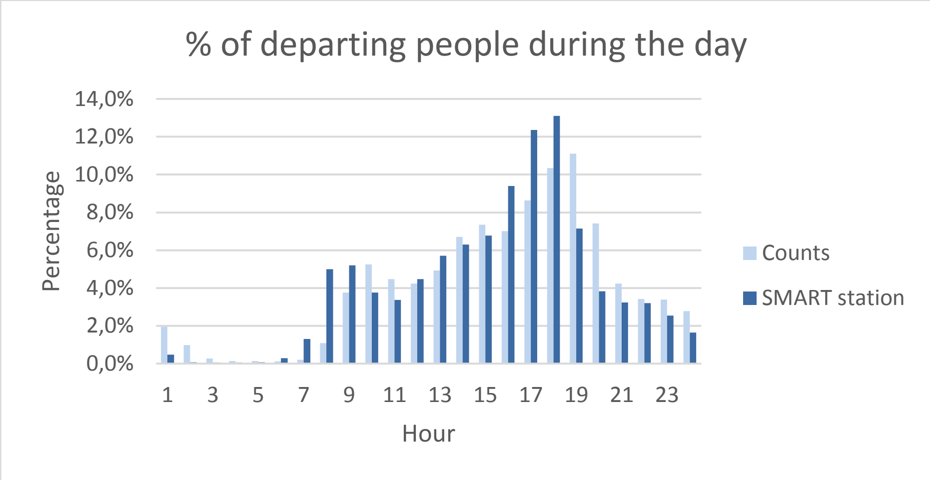

Figure 20: % of departing people during the day

TRAVELLING PEAK OR OFF-PEAK

The data from KIS can be used to validate whether the SMART station data is divided as expected over the peak hours and off peak hours (NS, 2011). Table 10 shows the spreading over each peak and the off-peak period. As shown above, the peak hours are slightly over represented. However, it is very close to the percentages indicated by the KIS data. A Kolmogorov-Smirnov test was performed on the distribution of both datasets over the time of day. The significance was 0.996 on a 95% confidence interval. This means that the distributions of KIS and SMART station over the time of day are equal.

Table 10: Travelling peak or off-peak

Time of the day	KIS data (NS,2011)	SMART station data
Morning peak	11%	14%
Evening peak	30%	33%
Off-peak	59%	53%

TIME SPENT IN THE STATION BUILDING

The time spent in the station building can be compared to previous research on the time spent in station buildings completed in the Netherlands (van Hagen, 2011). Data was collected by following people through the station building and using a stopwatch to measure the time spent. The method is very time consuming, so not a lot of observations were made (N=129). Four medium sized stations in the Netherlands were used for the observations: Enschede, Deventer, Zwolle and Amersfoort. This means that the data collected at Utrecht Central station is compared to medium sized stations (related to travellers per day), whereas Utrecht is a large station. It is expected that the time spent found at Utrecht will be higher than the time spent found by van Hagen (2011). He found that the mean time spent was 7:07 minutes with a standard deviation of 07:24 minutes. The distribution over time of the time spent at Utrecht Central station is shown in Figure 21. The mean time found is 10:31 minutes with a standard deviation of 12:28 minutes. The minimum is 45 seconds and the maximum is 2:59:47 hours. This is expected based on the research done by van Hagen (2011).

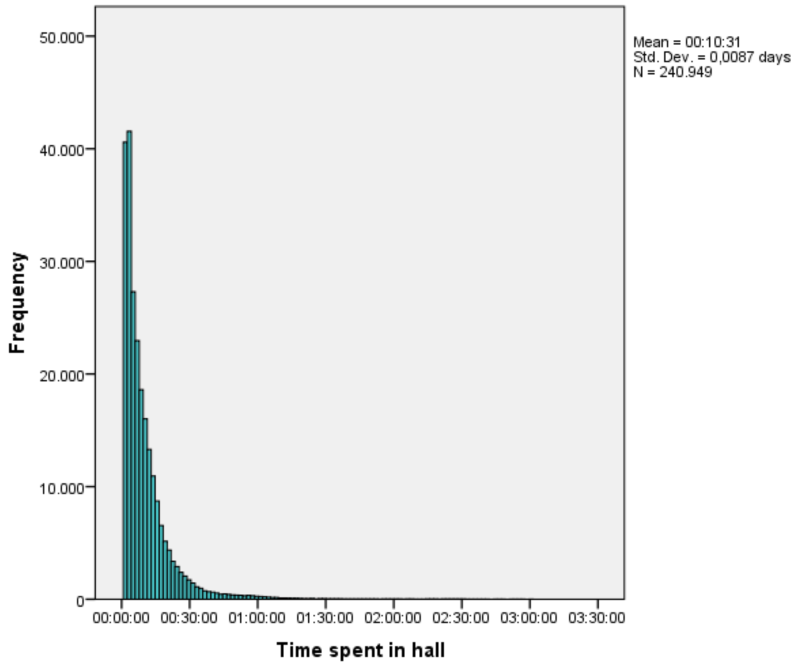

Figure 21: Time spent in station building - all departures

The maximum time spent in the station building is just below three hours. Due to the large number of departing passengers with a time spent around 10:00 minutes (+/- 50,000), the distribution around the maximum time spent in the station hall is not visible in Figure 21. Therefore it is interesting to zoom in into the section where the time spent is more than 1.5 hours. This is visualised in Figure 22. Up to 2.5 hours there are on average +/- 30 people in each 3 minutes section. However, after 2.5 hours the frequency drops severely. This means that the maximum time that pedestrians are willing to spent in the station is around 2.5 hours. Spending more time in the station hall seems to provide a lower gain for people, because nearly nobody stayed that long.

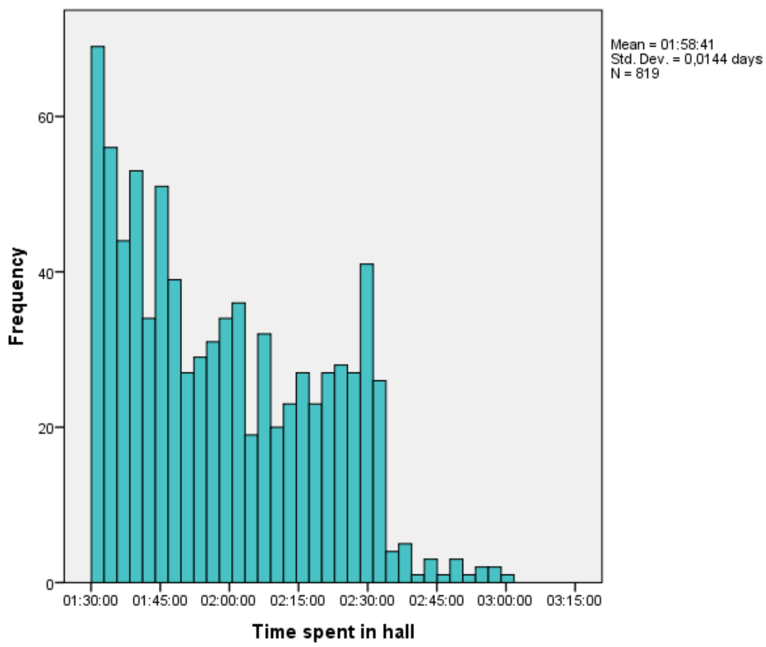

Figure 22: Time spent in station building - from 1.5 hours up

ACTIVITIES: SHOP VISITS

The shop visits are calculated based on the minimum time spent in the store (30 seconds) and the minimum signal strength of a smart phone. For every shop equipped with a Bluetooth and Wi-Fi scanner the total visitors measured is counted and presented in Table 11. This amount of shop visits adds up to a capture rate of 14.6%. The AH to go is the most popular shop in the station building according to SMART station data. This is also found in research on the shopping behaviour of train travellers done by NS Stations (NS Stations, 2013). Kiosk should have a higher number of visitors according to this research, but due to the malfunctioning of some of the scanners this was not found.

Table 11: Total amount of shop visits

Shop	Visits	% Visits of total
Rituals	334	0.9
Kiosk 5/7 N	0	0.0
Kiosk 8/9 Z	5	0.0
Kiosk 11/12 N	273	0.8
Kiosk 8/9 N	85	0.2
Paperchase	1,292	3.6
AH to Go (Etos)	15,009	41.9
AH to Go (8/9)	4,274	11.9
Broodzaak	3,111	8.7
Starbucks Bruna	3,140	8.8
T&S	4,299	12.0
Julia's	242	0.7
Burger King	1,697	4.7
Starbucks BK	803	2.2
Smullers	1,239	3.5
Total shop visits	35,803	100.0

DEPARTURES PER PLATFORM

The spreading of departing passengers over the platforms measured with SMART station is shown in Table 12. According to this data platforms 11-12 and 14-15 are most used. The trains on platform 11-12 depart in the direction of Amersfoort, which is one of the most frequently visited destinations from Utrecht Central station (NS, 2011). Therefore this number seems plausible for the total population. Platform 14-15 offers train services towards Nijmegen, Limburg and Rhenen. Neither of these belong to the top 5 of frequent destinations. Therefore it is expected that a slight overestimation of the departures on that platform is present. Platform 5-7 (trains for Amsterdam) should have a higher number of departures, because Amsterdam is the most visited station from Utrecht Central station (NS, 2011). Therefore, an underestimation of the departures from that platform are expected. This could also be due to many platform changes of the trains that are heading in the direction of Amsterdam (e.g. towards platform 11-12).

Table 12: Departures per platform

Platform	# departures	% departures of total
1-4	9,980	4.1
5/7	42,748	17.7
8/9	30,434	12.6
11/12	53,612	22.3
14/15	53,674	22.3
18/19	50,501	21.0
TOTAL	240,949	100.0

Several tests have been performed to check on the representativeness and validity of the data collected by means of SMART station. These tests showed that the distribution of travellers over the day and during the day (hourly and peak versus off-peak) are significantly equal to the data collected on the entire population (infrared scanners). The distribution of the time spent is as expected for Utrecht Central station. The shopping behaviour of the pedestrians in Utrecht Central station is partly as expected (e.g. AH to Go) and partly not (e.g. Kiosk). Therefore some issues related to the validity of this data could arise, especially related to shopping behaviour. It is important to take this into account in this research.

4.3.2. BLUETOOTH VERSUS WI-FI

The total dataset consists of data collected from Bluetooth and Wi-Fi devices. This dataset seems to provide a valid and representative view of the total population of departing passengers at Utrecht Central station related to the tests performed. Only the shopping behaviour of the pedestrians provided a skewed representation. A second test on the validity and representativeness of the data is to see at which points Bluetooth and Wi-Fi data show differences and comparisons. Each scanner placed in the station of Utrecht Central station makes use of a Wi-Fi and Bluetooth scanner. However, it is expected that different types of people make use of both technologies. First of all, there are much more people using Wi-Fi than Bluetooth: 91.5% versus 8.5%. The data on these technologies are tested on the same aspects as the total dataset: departures per day, departures during the day, travelling in peak hour or off-peak hours, time spent in the station building, shop visits and departure from platforms.

DEPARTURES PER DAY

As mentioned before, Bluetooth is less used than Wi-Fi by pedestrians. Therefore, a comparison in absolute numbers is difficult. Figure 23 shows the percentage of Bluetooth and Wi-Fi data per day compared to the total data collected. Except for 17-19 September 2013 no big differences are present in the data collected by each of the technologies compared to the total departures. There is an increase during the weeks in departing passengers, due to the ending of the school holidays and start of school and work. However, no explanation was found on why Bluetooth increases more than Wi-Fi. Apparently, relatively more people using Bluetooth were present in the station during the last week of the data collection.

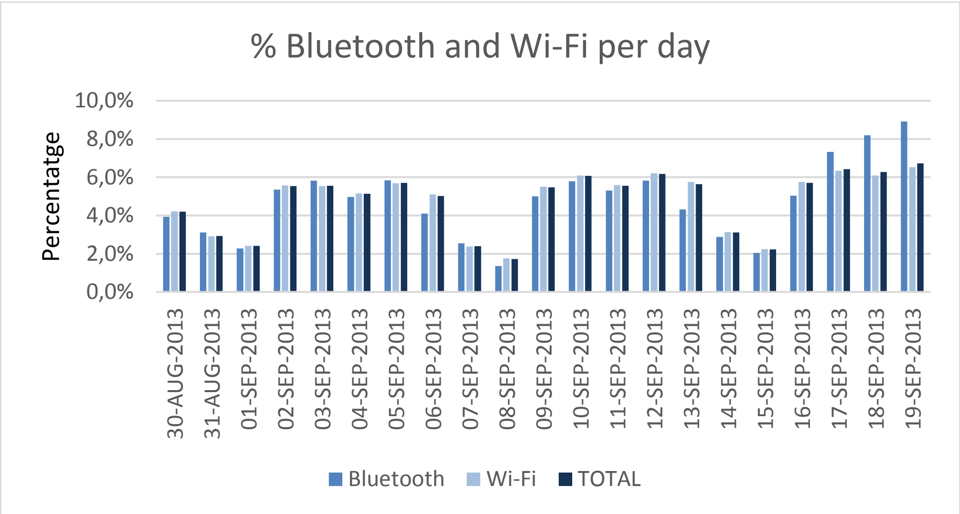

Figure 23: Departures per day - Bluetooth versus Wi-Fi

A Chi-square test on the relationship between the two technologies and the day of the data collection shows a value of 503.539. With 20 degrees of freedom the critical point is 31.4. The value shown is much larger, meaning that a relationship exists between the technology and the day of the data collection. This is confirmed by the Kolmogorov-Smirnov test, where the distribution of the data collected per day is tested against both technologies. The value is 0.591. On a 95% confidence interval this indicates that the distribution is the same for both technologies over the days.

DEPARTURES DURING THE DAY

When looking at the departing passengers during the day for both technologies, some interesting differences show up. Figure 24 shows the departures during the day. During the peak hours differences arise between the technologies. In the morning peak the share of Wi-Fi (on all Wi-Fi data) is higher and during the evening peak the share of Bluetooth (on all Bluetooth data) is higher. Bluetooth can be described as a wireless communication technology that can be used to connect devices with each other (Rogerson, 2014). A widely known application is the use of a wireless headset. This is very useful for people who make many work related phone calls. In general, this is less used in the morning when travelling to work. However, they might use it during the day and forget to turn it off or need it when returning home. This would explain the relatively small share in morning peak and relative large share in the afternoon peak. Another interesting difference arises at 15.00h. Here, Bluetooth is also over represented. This can be explained by the shift changes that happen for the train personnel of NS (Voskamp, 2012). At 15.00h the morning shift ends and the evening shift starts. NS personnel uses a device with Bluetooth enabled. A Chi-square test on the relationship between the technologies and the hour of the day provides a value of 1,451.455. The critical point with 23 degrees of freedom is 35.2. This indicates that a relationship exists between the technologies and the hour of the day. This is confirmed by the Kolmogorov-Smirnov test. The distribution across the hours of the day is the same for Wi-Fi and Bluetooth. The value is 0.893 on a 95% confidence interval.

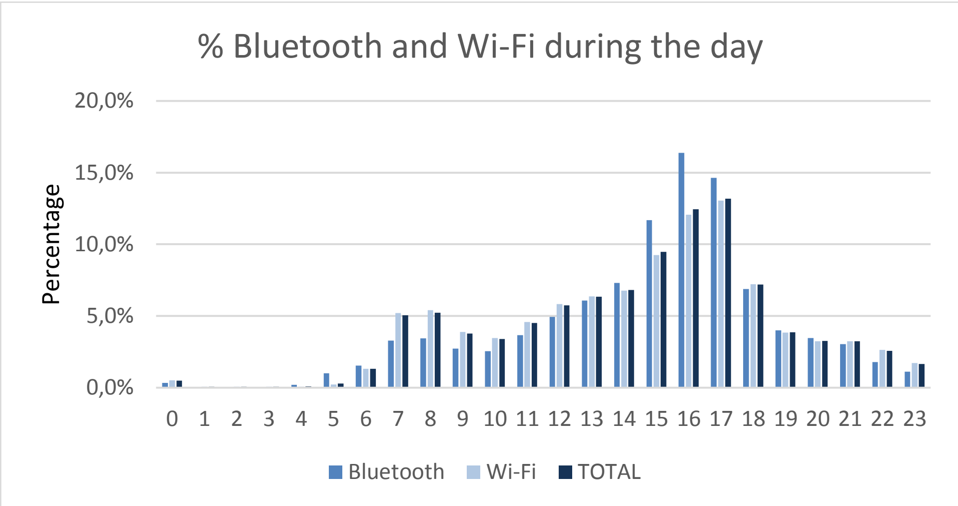

Figure 24: Departures during the day - Bluetooth versus Wi-Fi

TRAVELLING DURING PEAK HOURS OR OFF-PEAK HOURS

From the analysis on the departures during the day it was already clear that in the morning peak the use of Wi-Fi compared to the total use of Wi-Fi was higher than that of Bluetooth (compared to the total Bluetooth data). In the evening peak this was the other way around. Table 13 provides an overview of the peak hours versus the off-peak hours as a percentage of the total departures with that device. The Wi-Fi distribution is much more similar to the total dataset than the Bluetooth distribution. However, the differences between the shares are small. A Kolmogorov-Smirnov test shows that the distributions are the same. This means that in the distribution over the day is similar for Bluetooth and Wi-Fi (also related to the total).

Table 13: Peak versus off-peak travel - Bluetooth versus Wi-Fi

	Bluetooth	Wi-Fi	TOTAL
Morning peak	9.5%	14.5%	14.0%
Evening peak	37.9%	32.3%	33.0%
Off-peak	52.7%	53.2%	53.0%

TIME SPENT IN THE STATION BUILDING

The time spent in the station building can also be compared for Bluetooth and Wi-Fi data. Figure 25 shows the distribution of time spent for each technology (please note that the y-axis is not equal distributed). The peak for Bluetooth lies in the first 3 minutes, whereas the peak for Wi-Fi lies in the second 3 minutes (3-6 minutes). The average time spent is 10:36 minutes for Bluetooth and 9:42 minutes for Wi-Fi. This difference in average can be caused by the higher peak of the Bluetooth data. The standard deviation is 12:12 minutes for Wi-Fi and 14:58 minutes for Bluetooth. Minimum for both technologies is 45 seconds. The maximum for Wi-Fi is 2:59:47 hours compared to 2:56:28 hours for Bluetooth. For this case the relative inaccuracy of Wi-Fi compared to Bluetooth should be taken into account. The average inaccuracy of 10 to 20 seconds of Wi-Fi means (with a maximum of 45 seconds) probably causes the difference in the averages of both technologies. The actual averages probably lie closer to each other than given in this distribution. As already determined above, the distributions of the technologies over time spent are not equal. The Kolmogorov-Smirnov test shows that they are significantly different from each other with a 0.000 value on a 95% confidence interval.

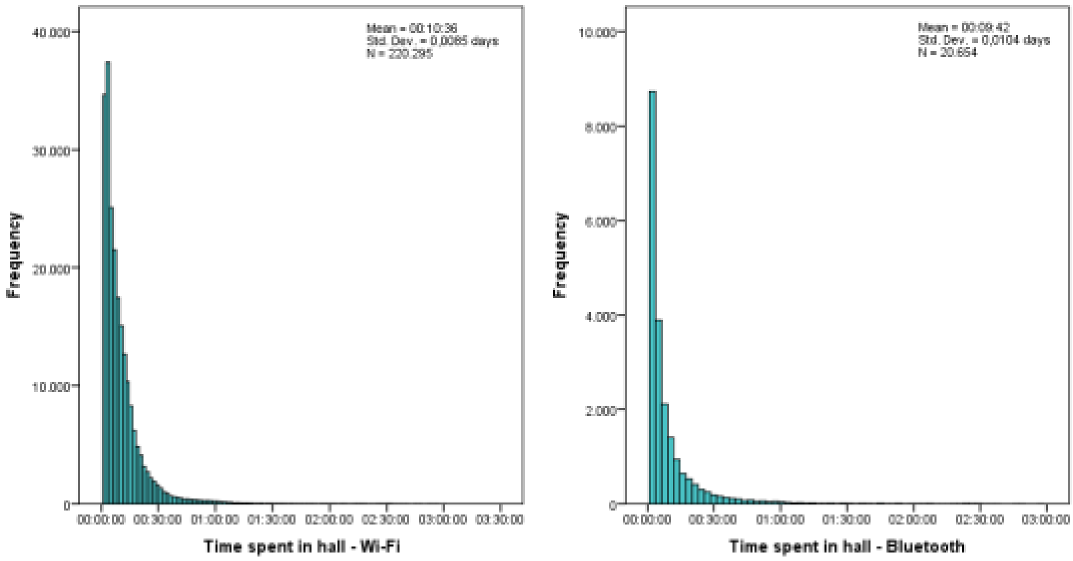

Figure 25: Time spent in station building - Bluetooth versus Wi-Fi

Since the tail of the distribution is not clearly visible, this is also enlarged. The departures over 1.5 hours are selected, just as in the validation of the total dataset. This is shown in Figure 26. In both cases a drop is visible after 2.5 hours. For the data on pedestrians that stayed in the station over 1.5 hours the distribution is similar for both technologies (according to the Kolmogorov-Smirnov test). The significance is 0.590, which is higher than the 0.05 belonging to the 95% confidence interval.

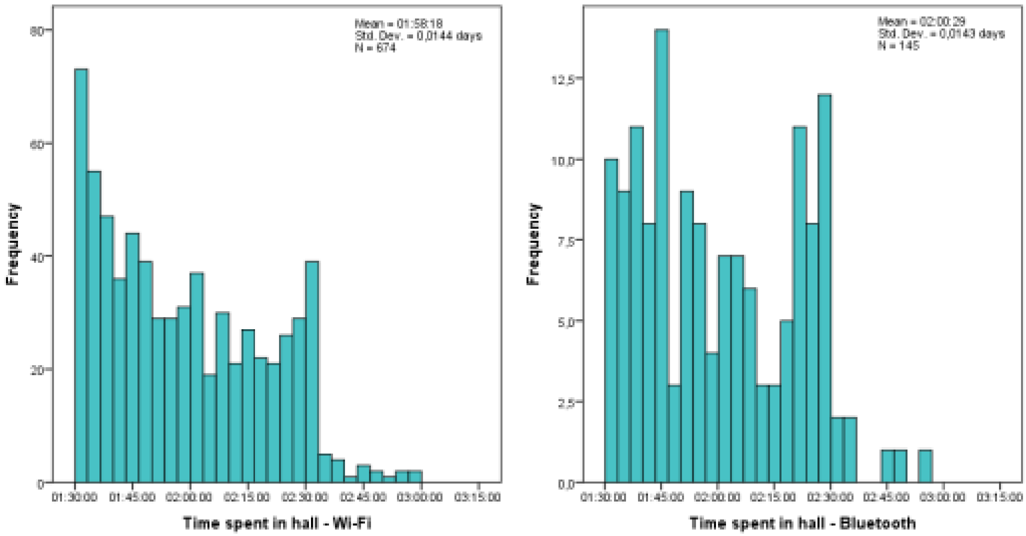

Figure 26: Time spent in station building >1.5h - Bluetooth versus Wi-Fi

ACTIVITIES: SHOP VISITS

Due to the large difference in the shares of Bluetooth and Wi-Fi it is not possible to look at the shop visits in absolute numbers. For a comparison the amount of shop visits per shop are showed as a percentage of the total shop visits. Figure 27 shows the percentage of shoppers that visit each shop.

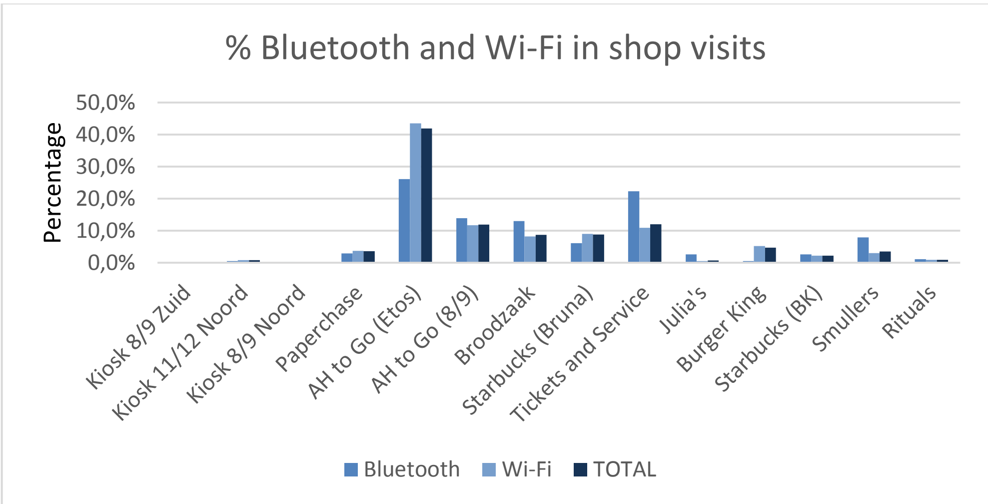

Figure 27: Visits per shop - Bluetooth versus Wi-Fi

As was already shown in the validation of the total dataset, the AH to Go (Etos) is the most frequently visited shop. Together with the AH to Go (8/9) it serves over half of all the departing passengers who have visited one or more shops. This is also visible when looking at both technologies. However, an interesting issue arises when looking at the AH to Go (Etos). The visits of that shop have a much higher share of the total for Wi-Fi users than for Bluetooth users (44% versus 26%), because this difference is very large. Wi-Fi overrules Bluetooth in number of users, this difference is also reflected in the total dataset. This could mean that an error was introduced in the definition of shopper for Wi-Fi compared to Bluetooth at the base of the SMART station method. On the other hand, the Tickets and Service shop shows a much higher visit number for Bluetooth users. The use of this shop over the day is relatively stable, meaning that it is visited in the peak hours but also in between the two peaks. This could mean that Bluetooth users more frequently buy tickets at this shop instead of the vending machines compared to Wi-Fi users. Or it could mean that NS personnel is also present in the Ticket and Service shop on regular occasions. The personnel uses Bluetooth devices, which can explain the skewed distribution in this shop. For the other shops the differences are much smaller. After performing a chi-square test on the distribution of Bluetooth and Wi-Fi over the shops it can be concluded that there is a relationship between the two technologies and the shop visits. The chi-square value is 1,229.471 where a critical value of 22.4 is introduced for 13 degrees of freedom. This value is larger, therefore the relationship is significant.

DEPARTURES PER PLATFORM

In the validation of the total dataset it was established that platform 11/12 showed the most representative numbers related to the most frequently visited stations from Utrecht Central station (NS, 2011). Platform 5/7 showed an underestimation and 14/15 a slight overestimation. In Table 14 the use per platform per technology is showed. Platform 1-4 is underrepresented by Wi-Fi users whereas platform 5/7 is overrepresented, compared to Bluetooth users. The difference in use per technology is smallest for platform 11/12, meaning that the relative share of that platform is nearly equal for both technologies. A chi-square test is performed to check on the relationship between the device used and the platform chosen. The resulting value is 631.591 which is higher than the critical value of 11.1 at 5 degrees of freedom. This means that there is a significant relationship between the device used and the platform used.

Table 14: Departures per platform - Bluetooth versus Wi-Fi

Platform	Bluetooth	Wi-Fi	TOTAL
1 - 4	6.1%	4.0%	4.1%
5/7	13.2%	18.2%	17.7%
8/9	13.9%	12.5%	12.6%
11/12	22.0%	22.3%	22.3%
14/15	25.5%	22.0%	22.3%
18/19	19.3%	21.1%	21.0%
TOTAL	100.0%	100.0%	100.0%

Several tests have been performed to check on the differences and comparisons between the Bluetooth and Wi-Fi data. These test showed that the distribution of departing pedestrians is equal over the day and during the day for both technologies. The time spent in the station building did differ for both groups. However, the distribution of the longer times spent is again equal. The distributions over the platforms and shops are again equal for both groups. However, in the AH to Go (Etos) the Wi-Fi users are overrepresented, whereas in the Tickets and Service shop the Bluetooth users are overrepresented. This means that related to the time spent and these specified shops issues concerning the validity could arise. It is important to take this into account in the analysis.

4.4. DISCRETE CHOICE MODELLING

The method used to perform the quantitative analysis on the data collected at Utrecht Central station was discrete choice modelling. This section provides a brief introduction to discrete choice modelling. The models estimated in this research aim to provide insight into the route and activity location choice behaviour of the departing pedestrians in the station building.

Decisions in the station, like route choice, are mutually exclusive. This means that only one option can be chosen (Ben-Akiva & Bierlaire, 1999). These reasons make discrete choice analysis is very suitable to analyse and predict decision making in the station building. As mentioned before the concept of utility will be used for the estimation of the models in this study. This concept is a decision rule that helps evaluate the different choices of a pedestrian in the station. There are many other decision rules that can be used to model the choice behaviour of pedestrians, however the concept of utility is widely used in transport related studies.

Therefore, the use of utility in this study seems valid. Examples of other decision rules are: dominance, satisfaction or lexicographic rules (ranking on importance) (Ben-Akiva & Lerman, 1985).

The discrete choice models developed in the state-of-the-art based on the concept of utility belong to logit and probit families. Logit models are widely used due to their flexibility and relatively low complexity compared to probit models. This research will therefore use logit models.

One use of utility which is often applied is utility maximisation (Ben-Akiva & Lerman, 1985). This concept assumes that the decision maker (the departing pedestrian) chooses the alternative that provides the highest subjective utility. Another example of the use of utility is regret minimization (e.g. Chorus et al., 2008). The use of utility maximisation is widely known and applied and will therefore be used in this study.

As mentioned before, the logit model is very flexible and easy to use. The basic model is the binary logit model. This model however imposes constraints related to the covariance structure (Ben-Akiva & Bierlaire, 1999). The result is that many other logit models have been developed that try to relax these restrictions. The models that can be used for modelling the route and activity location choice behaviour of departing pedestrians in the train station are briefly introduced below. The models concerned are multinomial logit, nested logit, cross-nested logit, path size logit and mixed logit[1,2]. A brief explanation on these models can be found in Appendix E. For more extensive information on the models the reader is referred to e.g. Train (2003), Hensher & Greene (2003) and Manski & McFadden (1981).

- ***Multinomial logit (MNL)*** → The multinomial logit model is the extension of the binary choice model, because it considers more than two alternatives. It assumes a linear and additive utility function. The model also assumes that the random error term is independent and identically Gumbel-distributed. The model has the property independence from irrelevant alternatives, which means that the relative probability that a departing pedestrian selects one alternative from a pair of alternatives is independent of any other alternatives. This basically means that no correlated alternatives can be present in this model. If there are, the model will not function properly.

- ***Nested logit (NL)*** → The nested logit model can handle correlated alternatives. The model divides the alternatives into nests. Within a nest correlated alternatives can be present, because the utility of a nest is based on the utility of each member of the nest. The random error term is assumed to be independent and identically Gumbel-distributed. The model does not allow for correlationship between the nests, meaning that only one parental nest can exist. If multiple nests seem to be present, the model will not function properly.

- ***Cross-nested logit (CNL)*** → The cross-nested logit model is a direct extension of the nested logit model, which resolves the issue of allowing only one nest. This model allows alternatives to be member of multiple nests.

- ***Mixed logit (ML)*** → The mixed logit model can be derived for many different behavioural specifications. Each derivation provides a different interpretation of the model. In this study the taste heterogeneity of the departing pedestrians is applied. The model shows the different weights

[1] Panel data can be considered because multiple pedestrians visit the station more than one time in the period of data collection. However, due to privacy reasons it is not known who visits the station on a more regular basis. Therefore, it is not possible to estimate this model.

[2] Path size logit model corrects the utility function to take overlap with other alternatives into account. It does not use nests like the nested and cross-nested logit model. However, the benefits concerning the model fit are limited compared to the multinomial mode. Therefore, it was chosen to explore nested and cross-nested logit model instead.

the departing pedestrians attach to each attribute. A different weight provided to an attribute identifies that different sub groups of pedestrians are present in the population.

In these models it is assumed that the each pedestrian is aware of the alternatives present in the station building. However, in Chapter 2 the distinction was made between familiar and unfamiliar pedestrians. The familiar pedestrians are expected to know all the alternatives presented in the station building. However, the unfamiliar pedestrians do not know/might not find all alternatives. This is represented in Table 15. The alternatives considered can differ for these groups. This cannot be captured by the model and should therefore be taken into account in the interpretation of the results.

Table 15: Choice set and alternatives (adapted from Hoogendoorn-Lanser (2005))

Choice set	Alternatives
Universal set	Existing alternatives
Subjective master set	Known alternatives
Subjective choice set	Feasible alternatives
Considered set	Considered alternatives
	Chosen alternative

4.5. CONCLUSIONS

This chapter provides insight into the research approach for the theory assessment. The case study location is Utrecht Central station. The train processes at the station are of major influence on the behaviour of departing pedestrians, because the train is the final destination within the station building. Therefore a proper understanding of these processes is necessary. The different platforms present at Utrecht Central station can be divided into three categories: final and start platforms, intercity platforms and mixed platforms. The first category is classified by the characteristic that (nearly) all the trains stopping there have Utrecht Central station as their final destination. On these platforms the dwell time is large. The second category is classified by the characteristic that only intercity services address that platform. This means that a rather constant behaviour can be expected on these platforms in regard to train services. The last category is classified by the characteristic that both intercity and sprinter services use the platform. This means that there is a larger variety of train services present. This might influence the behaviour of pedestrians on those platforms.

The data collected at Utrecht Central by means of SMART station needed to be prepared. The data comes from individual scans which need to be combined into a route for each pedestrian. Also, since the influence of the time table and train operations on the choice behaviour of departing pedestrians needs to be tested a match is made between the train process data and the SMART station data.

The SMART station data has been validated by means of several tests and comparisons in order to see whether the data is representative for the entire population of departing pedestrians in Utrecht Central station. It turns out that the pattern of departing pedestrians arriving in the station each day and during the day is similar to the population. The weekends are less busy than the weekdays and the afternoon peak is highest for departing passengers. Besides that the distribution of the time spent in the station building is as expected. Most pedestrians spend a relatively short time in the station building and some spend a longer time. When looking at the activities performed in the station building two things come to mind. The AH to Go is visited by nearly half of the pedestrians that visit a shop according to the SMART station data. It is expected that this shop is most popular, however the difference should not be this large. The Kiosk shops do have a lot of visitors as well. However, in the SMART station data this is not visible. This could be due to the fact that the scanners could not handle the large amount of travellers alighting the trains. Finally, most departing passengers are present on platforms 8/9, 11/12 and 14/15. This means that most departing pedestrians head in the direction of Amersfoort, Gouda and Nijmegen. However, according to data from the Netherlands Railways Amsterdam is the direction which is most frequently taken. This means that an under-representation of these pedestrians is present. In general, the SMART station data seems to be representative of the entire population of departing passengers.

The SMART station data is collected by means of Bluetooth and Wi-Fi scanners. Each scanner in the station building is equipped with both. It is expected that these technologies are used by different types of pedestrians. Bluetooth's most popular function is the wireless headset for making phone calls, whereas the main use of Wi-Fi is for accessing the internet. The first is therefore expected to be me more present amongst business people whereas the latter could cover a large group. Therefore the same validation tests and comparisons are done for these two groups as was done for the total dataset. The two technologies provide an equal image regarding the presence of the pedestrians over the days and during the day. However, it shows a different picture regarding the time spent in the station building. The Bluetooth user spends less time in the station building on average compared to the Wi-Fi user. Regarding the shop visits an interesting aspect arose. The Tickets and Service shop is visited more often in comparison to the other shops by people using Bluetooth. This could be due to the fact that NS service personnel use devices equipped with Bluetooth. NS personnel is thus present amongst the departing pedestrians. They are mainly registered when starting their shift. The number of NS personnel compared to the total dataset however is limited. Finally, regarding the platform of departure it can be said that on most platforms the distribution is equal. However, tracks 5 and 7 are more often used by Wi-Fi users compared to Bluetooth and for tracks 1-4 this is the other way around. In general, some differences between the users of the two technologies arise which need to be taken into account in the model estimation.

The dataset that is tested and validated will be used for the quantitative analysis of the theory. The method used for quantifying the relationships selected in Chapter 3 is discrete choice modelling. Four models are identified that can be used to asses the relationships selected for route choice and activity location choice: multinomial logit, nested logit, cross-nested logit and mixed logit. The route choice model that is estimated is addressed in Chapter 5 and the activity location choice model is addressed in Chapter 6.

5. ROUTE CHOICE MODEL

Part of the objective of this research is to understand and predict which factors and processes influence route choice behaviour of departing passengers in a station building. This chapter quantifies the selected relationships concerning route choice (Chapter 3) by means of a discrete choice model. The following question is addressed:

‘(5) Which factors have a significant influence on the route choice behaviour of departing pedestrians?’

In the framework a simultaneous decision between route choice and activity location choice was assumed for pedestrians familiar within the station building. In this case the route choice model is estimated on its own, because the influences on this individual decision are also interesting to know. A case in which the separation of these choices (related to route choice) is valid is where no activities are performed and thus no location needs to be chosen. This is the case when people head straight to the platform after entering the station building.

It is fixed beforehand which entrance a departing pedestrian takes (Jaarbeurs or Hoog Catharijne) and therefore it cannot be taken into account as part of the alternatives in the route choice model. In addition, the platform of departure is fixed. If someone would like to travel to Groningen they need to depart from platform 11/12. This aspect is therefore also excluded from the route choice model. The data collected with SMART station provides information on several locations of the route (binary). Therefore, the only real choice left is the stair choice. No data is available on the exact movements of the departing passengers, the only information regarding the route choice of the pedestrians heading straight to the platform comes from the stairs. This means that in this case the route choice could also be called stair choice. Please note that the entire route: entrance – stairs – platform is registered and used in the model.

Section 5.1 discusses the research questions relating to the route choice model. After that the platform chosen for this study is addressed in Section 5.2. Then the departing passengers of relevance are selected in Section 5.3. Section 5.4 discusses the factors and processes that will be taken into account in the model. There the definitions of each of the factors and processes are discussed. Also a first insight in the relationships is given. Section 5.5 addresses the results and findings of the route choice model. Finally, Section 5.6 discusses the conclusions of the route choice model.

5.1. RESEARCH QUESTIONS

Research questions need to be identified to provide a guide for the route choice model that is estimated. These questions are based on the theoretical framework and selected relationships. The selected factors and processes that are included in the route choice model are as follows:

- Time spent
- Travel time
- Time of day or week
- Distance
- Time table
- Train operations
- Visibility
- Orientation

In this route choice model no activities are performed, this means that the time spent in the station building and the travel time are the same.

Several factors need to be translated into operational attributes before they can be included into the model estimation. These are the time of day or week, time table and train operations. The time of day or week serves as a proxy variable for the familiarity of the pedestrians. Peak hours are reserved for familiar pedestrians whereas off-peak hours are reserved for unfamiliar pedestrians. There exists a relationship between the time of day or week of travel and the familiarity of the pedestrians, however this relationship can never be 1 on 1. Therefore, this proxy only serves as an approximation of the real relationship. This needs to be taken into account in the model estimation and interpretation. The time table is translated into the stop location of the train (whether the train stops underneath the stairs or not). The train operations are translated to the delay of the train.

The research questions relating to the factors to be tested can now be identified:

1. What is the influence of the *time spent/travel time* on the route choice?
2. What is the influence of the *peak hours versus off-peak hours* on the route choice?
3. What is the influence of the *distance* from entrance to platform on the route choice?
4. What is the influence of the *stop location of the train* on the route choice?
5. What is the influence of a *delay* of the train on the route choice?
6. What is the influence of the *orientation* on the route choice?
7. What is the influence of the *visibility* of the stair on the route choice?

5.2. PLATFORM SELECTED FOR ROUTE CHOICE MODEL

As was established in Chapter 4, there are some major differences between the platforms. Therefore, it was decided that one platform is used for the route choice model. A multi criteria analysis (MCA) was executed which helped select the best platform for the route choice model. The criteria by which the platforms were tested can be seen in Table 16. These criterion reflect both the importance of the train processes for this route choice model and the factors that will be researched in the model.

A weight is provided to each of the criteria. The time spent distribution and peak/off-peak need to be representative for the total population of departing passengers. Therefore, these criteria are provided a high weight. The distribution of trains over one hour is also awarded a high weight. This is done because if an unequal distribution is present (for example two trains depart from the same platform within a very short time) errors might be introduced by the matching algorithm used to match the train process and SMART station data. The other train process related criteria are less important than the distribution over the hour. The distance from entrance to platform is considered least important in the MCA.

Table 16: MCA criteria and weight - platform

Criterion	Best score?	Weight
Distance entrance - platform	Most variation	1
Distribution trains over one hour	Most equal distribution	4
Distribution of service types	Most uniform	2
Delay of trains	Most delay	3
Disrupted trains	Least disrupted trains	3
Time spent distribution versus SMART station	Most comparable	5
Peak/off-peak distribution versus SMART station	Most comparable	5

Six platforms are present at Utrecht Central station, however not all of these platforms can be included in the route choice model. Only one stairs provides access to tracks 1 to 4. This means that no stair choice is possible for this platform. The platform providing access to tracks 5 and 7 has two entrances. One of these serves as a double escalator for arriving passengers during peak hours. This means that no stair choice can be measured during peak hours. Therefore, these two platforms are excluded from the MCA and the route choice model. This means that the MCA is executed for four platforms.

The results of the MCA analysis are shown in Table 17. The analysis of each criteria can be found in Appendix F. The best platform according to the MCA has the highest score for the combination of criteria, which is platform 11/12. For each criteria a score of 1 up to 4 was provided, 4 points for the best scoring alternative on that criterion. This value was multiplied with the weight of the criterion on the total. Platform 11/12 scored best with 5 out of 7 criteria. For distance it scores badly, because the distances from both entrances to the platform are only slightly different. In addition, it has the second highest score for delay.

Table 17: Multi criteria analysis on platform selection

Platform	Distance	Distribution of trains	Distribution service type	Delay	Disruptions	Time spent	Peak/ off-peak	TOTAL
8 & 9	3	*16*	*8*	6	3	10	5	51
11 & 12	2	*16*	*8*	9	*12*	*20*	*20*	*87*
14 & 15	1	4	2	*12*	6	15	15	55
18 & 19	*4*	8	4	3	9	5	5	38

Platform 11/12 will be used for the estimation of the route choice model of departing pedestrians heading straight to the platform after entering the station building. The routes that are possible in the station for these pedestrians are shown in Figure 28. A total of 6 possible routes are drawn in the figure. However, because the entrance is determined in advance by departing pedestrians the decision on which route to take is determined by the stair choice: North, Mid or South. Figure 29 shows pictures of the stairs at platform 11/12.

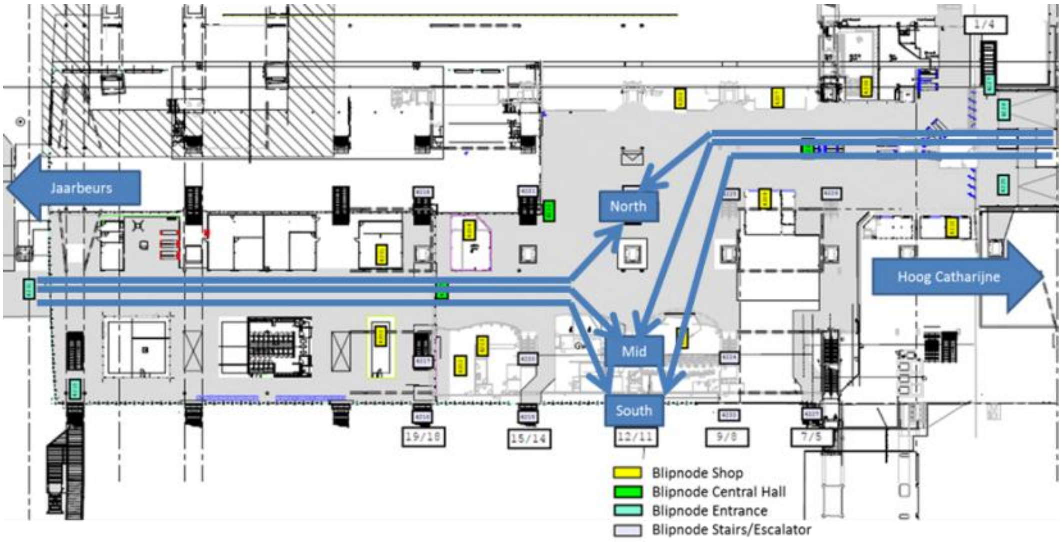

Figure 28: Visualisation of route choice on platform 11/12

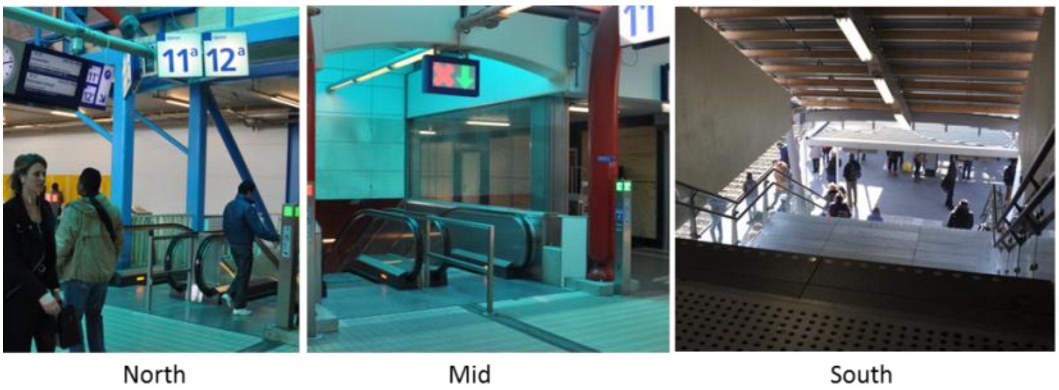

Figure 29: Pictures of the stairs on platform 11/12

5.3. SELECTION OF THE DEPARTING PEDESTRIANS

The total dataset that is collected by SMART station captures much more information than needed for the route choice model. Therefore a selection needs to be made on which departing pedestrians to include and which to exclude. The selection is based on multiple steps, the process is visualised in Figure 30.

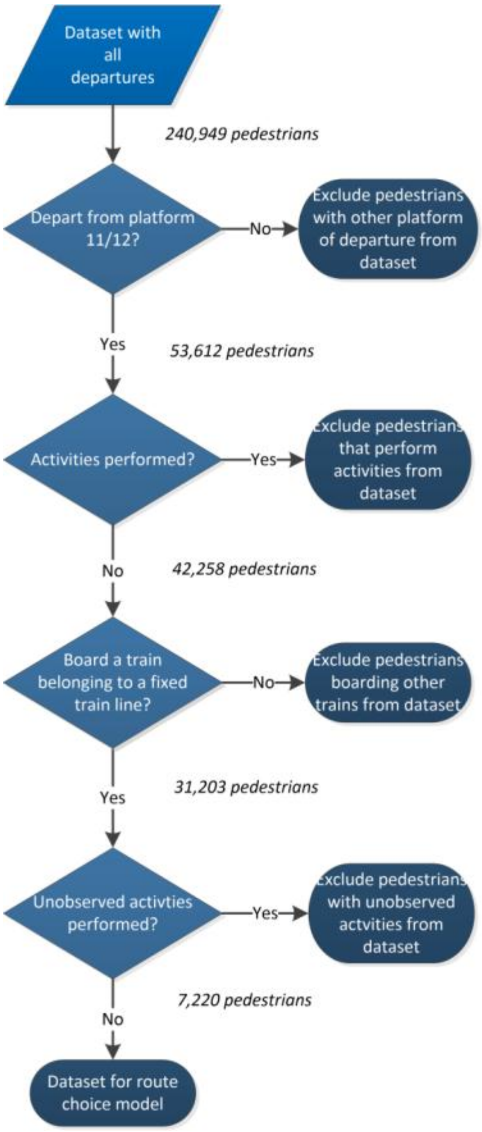

Figure 30: Data selection for the route choice model

The first step is to select only those pedestrians that depart form platform 11/12. This means that 53,612 pedestrians remain in the dataset. The route choice model is aimed at the pedestrians heading straight to the platform. Therefore the pedestrians performing activities (registered by SMART station) need to be removed. After this selection 42,258 pedestrians remain.

As mentioned before, the data of the train processes and SMART station are matched to be able to find out what the effects of the public transport factors are on the choice behaviour of pedestrians. Therefore, it is important to stick to the original schedule with only the fixed train lines departing from each platform. Other trains present on platform 11/12 were changed to this platform due to engineering works, delays or disruptions. This means that no stable situation is present for the data. Therefore, only those pedestrians that board a fixed train line from platform 11/12 are selected. After this selection 31,203 pedestrians remain in the dataset.

The final step is to exclude those pedestrians that performed unobserved activities. In Utrecht Central station approximately 55% of the shops are equipped with SMART Station. This means that in the other 45% activities are not registered by SMART station. In addition, people can wait in the station hall, which is not actively registered. This can only be seen by a higher time spent in the hall compared to walking straight to the platform.

In order to ensure that only pedestrians are included that walk straight from the entrance to the platform, a minimum walking speed can be used. The time spent from the entrance to the platform is known. The distance (minimum) from entrance to platform is known. This means that the speed of the pedestrian can be calculated using the following formula:

$$v = \frac{x}{t}$$

The average speed of pedestrians walking is usually assumed to be 5 km/h. However, in the station building another situation is present. People might need to look for their destination or they might not be in a hurry. In that case someone could have walked straight to the platform but did not manage 5 km/h. Elderly people usually have a lower speed of around 3 km/h (van Venrooij, 2011). This also seems a more plausible speed for the average departing pedestrian in the station building. The time a pedestrian needs for walking from the entrance to the stairs at 3 and 5 km/h is visualised in Table 18. The difference between these walking times shows the possibility for fast pedestrians to perform activities.

Table 18: Walk time entrance - platform at 3 km/h and 5 km/h

Route	Distance (m)	Time at 3 km/h	Time at 5 km/h	Difference 3 and 5
JB -N	135	2:43	1:37	1:06
JB-M	140	2:49	1:41	1:08
JB-S	155	3:08	1:51	1:17
HC-N	100	2:01	1:12	0:49
HC-M	115	2:19	1:23	0:56
HC-S	130	2:37	1:34	1:03

In the case of 3 km/h the elderly people, people with suitcases, people not in a hurry and people that need to look for their destination are included. This means that the probability for a type II error is expected to be small (see Table 19). The minimum extra time for performing an activity is 30 seconds plus the time needed to walk to the activity and back to the platform. This means that it is possible to perform an activity when the person is fast. This means that the probability for a type I error is expected to be higher than the type II error. However, the probability that this happens is small. If the limit would be set higher, the probability on a type II error would increase drastically. Therefore, 3 km/h seems a valid limit.

Table 19: Conclusions on data versus reality

		Reality	
		True	False
Measured	True	Correct	Type I error: False positive
	False	Type II error: False negative	Correct

Most of the pedestrians (approximately 76%) have a speed that is lower than 3 km/h. This means that a lot of unobserved activities are present at platform 11/12. SMART station registered around 11,000 people that perform activities. Around 16,000 people perform activities that are not registered (or are waiting in the station building). It is unknown what the distribution is between the unobserved activities and waiting in the hall. However, performing activities or waiting in the hall can be expected, as it is more interesting to wait for the train in the station hall than on the platform.

The distribution of pedestrians heading straight to the platform versus the pedestrians waiting in the station hall or performing activities differs per route. On most routes the latter group occupies around 70-80%. However, the pedestrians heading from either entrance to the South stairs of platform 11/12 have a more equal distribution for heading straight to the platform versus performing activities or waiting. It therefore seems that from the people that prefer the South stairs, relatively more pedestrians do not perform activities compared to the other stairs. Regarding the Mid and North stairs most people do perform activities and are therefore excluded from the dataset. This results in a relatively large share of the pedestrians choosing the South stairs over the Mid stairs.

After this selection only 7,220 pedestrians remain in the dataset of pedestrians that head straight to the platform. The distribution is 60% North, 8% Mid and 32% South. This is the final selection of departing passengers that are taken into account in the route choice model.

5.4. FACTORS AND PROCESSES INCLUDED IN THE MODEL

In Section 5.1 the operational attributes that are included in the model were identified. These were time spent in the station building/travel time, peak/off-peak, distance, stop location of the train, delay, visibility and orientation. The comparison between Bluetooth and Wi-Fi users (Section 4.3.2) indicated that they have a different time spent. Therefore a check is completed on the distribution of time spent for this route choice model. This section addresses the factors as they are used for the model development. Also, a first analysis on the relationship between the factors and processes and route choice is provided.

TIME SPENT IN THE STATION BUILDING/TRAVEL TIME

In order to apply discrete choice analysis it is necessary to know which alternatives are included and what the characteristics are of each alternative. In this case the pedestrian can choose between three alternatives. A pedestrian has a different time spent or travel time towards each of these alternatives. The data collected by means of SMART station only shows the travel time towards the chosen alternative. This means that a calculation of the non-chosen alternatives is necessary. This calculation is based on the speed of the pedestrian over the chosen route (based on travel time and distance). The assumption is that a pedestrian walks towards the other alternatives with the same speed as he did to the chosen alternative. This assumption is only valid if no queues appear on the other alternatives. The bottleneck on these routes would be the escalator or stairs. However, the capacity of the stairs and escalators is so large that no significant queues arose during the day. Also, the assumption does not hold if the pedestrian spends nearly no time on the platform because the train departs. If he would have to go via an alternative further away, he would miss the train. This happens in 0.55% of the cases. Therefore the assumption seems plausible.

The formula applied when the North stairs is chosen and the travel time towards the Mid stairs needs to be calculated is as follows:

$$T_m = \frac{D_m}{\frac{D_n}{T}}$$

With T_m being the travel time towards Mid, D_m being the distance from the entrance towards Mid, D_n being the distance from entrance towards North and T the actual travel time. The formula is adapted for finding the other relationships. The unit of travel time is minutes.

The distribution of actual travel time is shown in Figure 31. The total number of observations per stairs differ, therefore the bars are higher for the North and South stairs than for the Mid stairs. An interesting observation from this figure is the visibility of peaks. At around 1:30 minutes a peak arises in the number of observations for all the stairs. This means that of this dataset people have a high preferences for being in the station building for around 1:30 minutes. A second peak is visible at around 2:15 minutes, however this peak is lower.

The average actual travel time also differs per stairs. Mid has the highest travel time and North the lowest. The average travel time for Mid is 1:59 minutes, for North is 1:40 minutes and for South is 1:49 minutes. The South stairs are the farthest away. This means that the average speed of the pedestrians taking these stairs is higher than the Mid stairs. The standard deviation also differs per alternative. Mid has the lowest standard deviation with 0:23 minutes. This means that the spread in travel time is limited. The standard deviation of North is 0:24 minutes and for South is 0:34 minutes. The average actual travel time to the South stairs was lower than from Mid, but the spread is also larger. This can indicate that this factor indeed influences the stair choice.

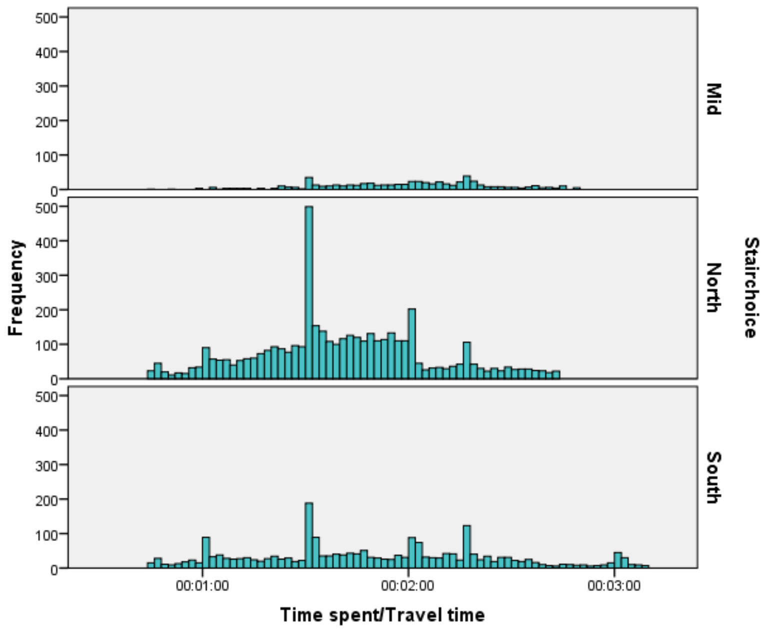

Figure 31: Travel time distribution over stair choice

A comparison between Wi-Fi and Bluetooth users for the time spent per stairs indicates that no differences arise between these user groups. This is shown in Table 20. The mean time has no significant difference on a 95% confidence interval. This means that the differences indicated in Section 4.3.2 do not hold for the pedestrians heading straight to the platform. Bluetooth and Wi-Fi users can be combined in the dataset.

Table 20: Bluetooth and Wi-Fi distribution over stair choice

Stairs	BT – mean	BT – std. dev.	Wi-Fi – mean	Wi-Fi – std. dev.
North	1:42 min	0:23 min	1:39 min	0:25 min
Mid	1:57 min	0:23 min	1:59 min	0:24 min
South	1:50 min	0:37 min	1:49 min	0:34 min

PEAK/OFF-PEAK

A pedestrian travels during peak hours or during off-peak hours. This factor is a proxy for the familiarity of pedestrians. This factor is not directly related to the alternatives of the route, it is related to demographics. It shows a distinction between pedestrians that travel during different moments of the day. This means that this alternative is not specific, but generic for all alternatives. The coding for this factor is done using a dummy variable. Travelling during peak hours is given a 1, travelling during off-peak hours is given a 0. Dummy coding is used because the effects of this variable on other factors that influence route choice need to be tested. By using dummy coding the interaction can be modelled in the most simple way.

The distribution of peak and off-peak over the stairs is shown in Table 21. Nearly no variation is present in the distribution over the stairs during both periods. This can indicate that the influence of this factor on route choice is much smaller than expected.

Table 21: Peak/off-peak distribution over stair choice

Stairs	Off-peak	Peak
North	1,946 (60.0%)	2,398 (60.3%)
Mid	216 (6.7%)	375 (9.4%)
South	1,080(33.3%)	1,205 (30.3%)

DISTANCE

The exact distance a pedestrian covers is not registered by SMART station. Therefore, the only distance known is the minimum distance of the route (measured from a map). The factor is alternative specific, since it differs per alternative. The distances in meters from entrance to stairs are shown below.

	North	Mid	South
JB	135	140	155
HC	100	115	130

The distribution of the distances in the dataset is represented in Table 22. The distribution of the stair choice differs greatly for the different entrances. The preference of pedestrians coming from Hoog Catharijne is strongly towards the North stairs, whereas the preference of the pedestrians coming from Jaarbeurs is more spread amongst North and South. This can indicate that the factor distance has an influence on the route choice behaviour of departing pedestrians towards platform 11/12.

Table 22: Distribution of distance over stair choice

Stairs	HC	JB
North	3,172 (68.0%)	1,172 (45.9%)
Mid	239 (5.1%)	352 (13.7%)
South	1,253 (26.9%)	1,032 (40.3%)

STOP LOCATION OF THE TRAIN

The stop location of the train can be obtained from the train process data. At Utrecht Central station the platform is divided into two phases: A and B. The train can stop on the A phase, B phase or cover both phases. The North stairs are located at the A phase, whereas the Mid and South stairs are located at the B phase. If the stairs provide direct access to the train, the alternative is given a 1. If this is not the case, it is given a -1. This type of coding is referred to as effect coding. The distribution of phases over the stairs is visualised in Table 23.

Table 23: Distribution of stop location of the train over stair choice

Stairs	Phase A	Phase B	Phase A & B
North	965 (66.5%)	207 (49.8%)	3,172 (59.3%)
Mid	77 (5.3%)	44 (10.6%)	470 (8.8%)
South	410 (28.2%)	165 (39.7%)	1,710 (32.0%)

The distribution differs per phase. The North stairs are more attractive if the train actually stops on the A phase. However, it is still very attractive when the train does not stop on that phase. The B phase is accessed via stairs Mid and South, combined they are chosen in 40-50% of the cases when the train stops on that phase, whereas they are chosen only in +/-30% of the cases when the train does not stop on the B phase. Therefore it is expected that this factor does influence stair choice.

DELAY

The delay of a train can be obtained from the train process data. A train is delayed according to NS standards when it departs more than 3 minutes later than scheduled. This standard is also used in this study. The delay of the train is assumed equal for each alternative, meaning that is generic. This factor is also coded using effect coding. This means that a case is given a 1 if there is a delay present and a -1 if there is no delay present. The distribution of having a delay or not over the stair choice is shown in Table 24. There is a small difference in the distribution over the North and South stairs. However the difference is not that large. Therefore the impact of this factor on stair choice is expected to be limited.

Table 24: Distribution of delay over stair choice

Stairs	Delay	No Delay
North	411 (53.3%)	3,933 (61.0%)
Mid	63 (8.2%)	528 (8.2%)
South	297 (38.5%)	1,988 (30.8%)

VISIBILITY

The visibility of the alternatives is related to the location dimensions. An alternative is either visible or invisible. This does not change over time (maybe in the long term when the station lay out is changed). The North and Mid stairs are visible, the South stairs is invisible. This means that it is specific per alternative, but it also shows no variation per pedestrian. This factor is also coded with effect coding. A visible stair is given a 1, a non-visible stair is given a -1.

The distribution of the visibility is therefore equal to the choice of the stairs. The North stairs are chosen in 60.2% of the cases. The Mid stairs are chosen in 8.2% of the cases and the South stairs are chosen in 31.6% of the cases. Due to lack of variability, the influence of this factor is expected to be limited.

ORIENTATION

The orientation reflects to the location of the stairs regarding the entrance (left versus right). This means that it differs per entrance used. The entrances are located at opposite sites of the station, therefore the orientation is mirrored. This factor is again coded using effect coding. The right side is given a 1 and the left side is given a -1. The reflection of this coding per route is shown below.

	North	Mid	South
HC	1	-1	-1
JB	-1	1	1

The distribution of this variable over the stair choice is visualised in Table 25. It is visible that the distribution of the actual orientation of pedestrians (that of the chosen alternative) differs greatly. Especially related to the North and South stairs. The South stairs are more frequently visited on the left side than on the right side. Regarding the North stairs, this is the other way around. It is therefore expected that this factor does influence the stair choice.

Table 25: Distribution of orientation over stair choice

Stairs	Left	Right
North	1,172 (44.0%)	3,172 (69.6%)
Mid	239 (9.0%)	352 (7.7%)
South	1,253 (47.0%)	1,032 (22.7%)

In summary, the analysis shows that a large influence on stair choice is expected from several factors. These are distance, time spent or travel time, train stop location and the orientation. The other factors (peak/off-peak, delay and visibility) did not show much variation over the stair choice.

5.5. RESULTS AND FINDINGS

In this section the route choice model is estimated using the software package BIOGEME (Bierlaire, 2003). This is done in several steps. As mentioned in Section 4.4, the multinomial logit model is the most basic model that takes more than 2 alternatives into account. Therefore this model is the starting point in the model estimation. The nested logit model, cross-nested logit model and mixed logit model are estimated as well because they might improve the MNL model estimated. Therefore, they use the best functioning multinomial logit model as input. In the estimation of the multinomial logit model, the following steps are taken:

1. Estimate a model for each of the factors and processes individually
2. Combine the significant factors and processes in a stepwise manner to create the best model

The first step is to create a model for each of the factors and processes. In each of these basic models alternative specific constants (ASC) are used that can capture all unobserved attributes of the model and show the base preference of the departing pedestrians towards the alternatives. An extensive overview of the model results can be found in Appendix G. Table 26 shows an overview of the parameters estimated for each individual model with the value, t-test, p-value, log-likelihood and adjusted rho-square value.

Table 26: Overview of the individual models of each factor and process - parameters

Parameter	Value	Robust t-test	p-value	Log-likelihood	Adjusted rho-square
Time spent/ travel time	-2.18	-33.10	0.00*	-6,069.938	0.234
Peak	$3.56e^{-15}$	0.00	1.00	-6,315.028	0.203
Off-peak	$2.52e^{-15}$	0.00	1.00		
Distance	-0.0269	-27.51	0.00*	-6,234.394	0.225
Stop location of the train	0.166	6.60	0.00*	-6,292.858	0.206
Delay	$-2.42e^{-15}$	-0.00	1.00	-6,315.028	0.203
Visibility	$1.06e^{-15}$	0.00	1.00	-6,315.028	0.203
Orientation	0.230	18.19	0.00*	-6,147.031	0.225

*Significant on a 95% confidence interval

Three factors turn out to provide no significant contribution to the utility of the alternatives. They are not different from zero. These factors are peak/off-peak, delay and visibility. This means that three of the research questions identified can already be answered. The peak and off-peak were coded as dummy variables, which means that two parameters need to be estimated: one to show the influence of peak and the other for the influence of off-peak. The peak and off-peak do both not influence stair choice of departing pedestrians. This means that the proxy time of day does not show a significant influence on the route choice and thus regarding this proxy there is no difference in the choice behaviour of familiar and unfamiliar pedestrians. However, this relationship is not 1 on 1, therefore it cannot be concluded that the familiarity of pedestrians does not have influence on the route choice. In addition, the delay of the train does not influence the stair choice of the departing pedestrian. People choose their stairs in the same way regardless of whether the train is delayed or not. Finally, the visibility does not influence the stair choice. Apparently departing pedestrians know of its location (as it is chosen relatively often).

Four factors are significant on a 95% confidence interval: time spent, distance, stop location of the train and orientation. The stop location of the train adds utility to the alternative. If the stairs provide direct access to the train this results in a bonus of 0.166 utility points. If not, these utility points are subtracted. The orientation provides a positive value to the utility function if the alternative is located on the right side. This was as expected, since people in the Netherlands are mainly right-oriented. The time spent or travel time provides added value to the utility function of the stair choice. The higher the time spent, the lower the utility for that alternative. Since the time spent increases as the distance is larger, the South stairs have a higher time spent or travel time than the Mid stairs, and Mid have a higher time spent than the North stairs. Since the distance is used to calculate the time spent a correlation is present between these factors. This correlation is significant on a 99% confidence interval. A greater distance is therefore also disliked by departing pedestrians. This means that these factors cannot be combined into one model.

As mentioned before, the model assumes that everyone considers these three alternatives. It might be that some of the pedestrians (e.g. unfamiliar pedestrians) do not consider all the stairs. This cannot be seen in the model. This limitation needs to be taken into account when deriving conclusions from the model.

The second step is to combine the significant attributes into one model. The distance and time spent cannot be combined. Since time spent has a higher added value to the model than distance, time spent is used. The best model is a combination of the three remaining factors: time spent, train stop location and orientation. The utility function for alternative *i* is then the following:

$$U_i = ASC_i + \beta_{ts} \times Time\ spent_i + \beta_{tr} \times Train\ stop_i + \beta_o \times Orientation_i$$

As mentioned before the South stairs have a higher time spent than the North stairs. This means that these stairs are always valued more negative related to the time spent than the North stairs. However, the South stairs are chosen relatively often (+/- 31%) compared to the Mid stairs (+/- 8%). Since the time spent is the most dominant factor in the model, it has a large impact on the utility of the South stairs in the model. This needs to be compensated in the shape of an ASC. This also results in a high correlation of the ASC North with the time spent parameter since it now basically provides a double negative value compared to the South stairs. This means that they explain the same utility. Therefore, it is decided that the ASC North should be removed from the model. This way the correlation issue does not arise.

The general information of the route choice model is shown in Table 27. The model can be tested on its performance using the likelihood ratio test. This value is provided in the table for the comparison of this model with the null situation. The model has 4 degrees of freedom and the critical value on 95% confidence interval is 9.49. The model scores a value of 3881.138 and is therefore significantly better than the null situation. A comparison of this model with the best individual model is also possible. This was the model on time spent or travel time. The following formula is used for the estimation of the likelihood ratio:

$$-2\ (LL2 - LL1) > \chi^2$$

The likelihood of the best individual model is -6,069.938 (2 parameters) compared to the likelihood of -5991.412 for this model. This means that the resulting likelihood ratio is 157.052. This is higher than 5.99 on a 95% confidence interval. Therefore the combined model is significantly better than the individual model. The adjusted rho-square of this model is also improved compared to the individual models.

Table 27: General information on the combined MNL model

Model	Multinomial Logit
Number of estimated parameters	4
Number of observations	7220
Number of individuals	7220
Null log-likelihood	-7931.981
Final log-likelihood	-5991.412
Likelihood ratio test (null situation)	3881.138
Likelihood ratio test (individual model)	157.052
Adjusted rho-square	0.244
Diagnostic	Convergence reached...
Iterations	6

The utility parameters that are estimated for the combined model are provided in Table 28. The ASC for South is fixed to zero. As mentioned before the North ASC is removed from the model. South and North therefore have the same starting point or base preference in this model. This is not entirely true given the choice distribution of pedestrians (60.2% North versus 31.6% South), however due to the dominance of the time spent this issue is resolved. All attributes are significant on the 95% confidence interval. The orientation is positive for the right side and negative for the left side. The direct access to the train provides positive utility whereas indirect access provides negative utility. A higher time spent means less utility for that alternative. This means that the sign of these parameters is as expected.

Table 28: Utility parameters for combined MNL model

Name	Value	Robust Std err	Robust t-test	p-value
ASC_M	-1.64	0.0436	-37.53	0.00
ASC_S	0.00	fixed		
Beta_o	0.156	0.0135	11.61	0.00
Beta_tr	0.145	0.0261	5.56	0.00
Beta_ts	-1.76	0.0736	-23.89	0.00

The orientation and train access from the stairs are coded with effect coding, meaning that they have the value 1 or -1. The orientation parameter has a higher value, implying that this factor is more important in the choice of a pedestrian than the direct access from the train. The time spent by a pedestrian ranges from 0.58 to 3.13. The value of this parameter is much higher than the other two. This means that this factor is more dominant than the other two. An example of a departing passenger is used to illustrate this.

Departing passenger A chooses the South stairs. The train stops on the A-phase and this person comes from Hoog Catharijne. The following factor values are found for this pedestrian:

- Orientation → North = 1, Mid = -1 and South = -1
- Train stop → North = 1, Mid = -1 and South = -1
- Time spent → North = 0.7, Mid = 0.8 and South = 0.9

The impact of the factors on the total utility of pedestrian A is the following:

U_{north} = 0.156 + 0.145 – 1.232 = -0.931

U_{mid} = -1.64 -0.156 – 0.145 – 1.408 = -3.349

U_{south} = -0.156 – 0.145 – 1.584 = -1.885

Pedestrian A has a probability of 26.1% for choosing South, 67.8% of choosing North and 6.0% of choosing Mid. Pedestrian A chooses South.

The example shows that the impact of the time spent is highest and therefore most important in the decision making. This can be reflected by the range of impact from each attribute on the stair choice. The impact of orientation is either 0.156 or -0.156, the impact of the train stop is 0.145 or -0.145 and the impact of the time spent is between 1.02 and 5.51. The time spent therefore has between 6 and 35 times more impact than the orientation and between 7 and 38 times more impact than the train stop.

The remaining four research questions can now be answered. The time spent is the most dominant variable that influences the stair choice of the pedestrians. Due to a significant correlation the distance was not taken into account in the combined model, however it does influence the choice (when modelled individually). The impact for this attribute ranges from 2.69 to 4.17, which implies that the impact on the choice behaviour is rather large (given that the ASC for Mid is -1.71). The orientation and train stop also influence the stair choice of pedestrians, but are of less importance.

Two nested models are estimated, that could improve the combined MNL model. The first is North versus South alternatives (Mid is combined with South). These alternatives have attributes that have a lot in common. It provides a significant improvement on the combined MNL model. However, due to high correlation amongst the parameters it is impossible to gain a proper interpretation of the model.

The second model is escalators versus stairs. This is essentially the same as visible versus non-visible alternatives. The North and Mid stairs are thus combined into a nest. This nest is expected because of the difference in the infrastructure of the alternatives. The model provides a significant nest, however the nest parameter is equal to 1. This means that it is the same model as the combined MNL model. The combined MNL model is simple and should therefore be used.

The presence of two nests in the model indicates that the cross-nested model might provide better results. The model indeed provides a significant improvement. Mathematically the model is correct, however the outcomes are not correct due to wrong signs in the parameters.

The mixed logit is not expected to improve the combined MNL model as the selection of the departing pedestrians made sure that the same type of pedestrians was selected for the model estimation. The estimation of the mixed logit model for each of the attributes indeed resulted in a worse model. The departing pedestrians in the model are a homogeneous group. Therefore, the combined MNL model is the best model.

5.6. CONCLUSIONS

Route choice consists of the entire route from entering the station building to boarding the train. For this route choice model only the departing pedestrians heading straight to the platform are taken into account. If no activity is performed, the route choice can be estimated separately from activity location choice. The platform of departure is already fixed on before hand (the train in a certain direction usually departs at the same platform), just as the entrance via which a pedestrian enters the station building. Given the limitations of the data collection method regarding the exact movements of the pedestrian, the only choice left that can be measured is the stair choice. The route then becomes entrance – stairs – platform, with entrance and platform fixed. The platform used for the estimation of the route choice model is 11/12. This platform was identified to be best suitable by means of a multi criteria analysis.

The factors and processes possibly influencing route choice selected in Chapter 3 are quantified by means of a discrete choice model. These factors are time spent, time of day or week, distance from entrance to stairs, travel time, time table, train operations, visibility and orientation. The time spent and travel time are considered the same for this model, as no activities are performed. The time of day or week is reflected by peak hours versus off-peak hours, the time table is reflected by the stop location of the train and the train operations are reflected by the delay.

Based on a first analysis on the relationship between the factors and processes and the stair choice it is expected that the time spent/travel time (no differences were found between Wi-Fi and Bluetooth users), distance, orientation and stop location of the train are of influence on the stair choice of the departing pedestrians. This was confirmed in the discrete choice model that is estimated. A combined model could be estimated taking travel time, orientation and the stop location of the train into account. The travel time and distance are highly correlated and can therefore not be combined into one model. As the travel time is the most dominant factor this was preferred over distance. In the estimated model time spent is most influential on the choice behaviour of pedestrians. If the travel time of a departing pedestrian is low (0:45 min) the impact of the travel time is 6 a 7 times as high as the factors orientation and the stop location of the train. If the travel time on the other hand is large (3:00 minutes), the impact can be as much as 35 a 38 times as high as the orientation and stop location of the train. The travel time is related to the distance covered. In order to ensure that no unobserved activities are performed, a minimum speed limit is established. This was determined to be 3 km/h. Therefore the range between travel times is not that large. But it can be concluded that pedestrians optimise their travel time in choosing a route. This was also found in literature.

If the stairs provide direct access to the train, this will make the choice for that stairs more likely. This can be explained by looking at the dynamic information signs above the stairs. If the train departs from the A-phase or B-phase, the information signs will only show the train information above the stairs that provide direct access. In that case the other stairs will show information for another train or are left blanc. Figure 32 provides an example for the information. This factor however is least important in the decision process. This can indicate that people know where their train departs or are only focussed on the platform and not the phase of departure. This significance of this factor provides new knowledge about the choice behaviour of departing pedestrians regarding their route choice.

Figure 32: Information provided above the stairs

Next to that, the stairs located on the right side of the departing pedestrian has a higher probability of being chosen. This can be explained based on two aspects. First of all, in the Netherlands everything is aimed at the right side: e.g cars drive right, bicyclist ride right and trains drive on the right. This means that the primarily orientation of people in the Netherlands is aimed at the right. This is reflected in the second aspect, the pedestrian flows in the station building. The flows at Utrecht Central station are mainly oriented right. Therefore, it is more logical to choose the right stairs. The flow needs to be crossed when choosing the left stairs.

The peak and off-peak hours do not influence the choice of stairs. In peak hours it is more crowded in the station building, but departing pedestrians choose their stairs regardless of the crowdedness and time of day. The peak and off-peak hours served as a proxy for the familiarity of the departing pedestrians. This means that regarding this proxy familiar and unfamiliar pedestrians choose their route/stairs the same way. The relationship between the proxy and familiarity however is not one on one, therefore it cannot be concluded that the familiarity of a pedestrian is of no influence on route choice.

The fact whether a train is delayed does not influence the stair choice of the pedestrian. Next to that, the visibility of the stairs are of no influence on the stair choice. One of the stairs is not visible to the pedestrian. It can be expected that a pedestrian unfamiliar with the station will not choose this alternative. It could be that he does not even consider this alternative. In the discrete choice model however it is assumed that all pedestrians consider the same set of alternatives. This issue regarding visibility and familiarity can therefore not be captured in the model.

6. ACTIVITY LOCATION CHOICE MODEL

Part of the objective of this study is to understand and predict which factors and processes influence activity location choice behaviour of departing passengers in a station building. These factors and processes are identified in Chapter 2. The selected factors and processes concerning activity location choice (Chapter 3) are quantified in this Chapter by means of a discrete choice model. The following question is addressed in this chapter:

‘(6) Which factors have a significant influence on the activity location choice behaviour of departing pedestrians?’

In the framework identified in Chapter 2, the relationship between route choice and activity location choice was assumed to be simultaneous for familiar pedestrians. The route choice model in Chapter 5 was estimated separately. The location choice is estimated alone. This means that in the activity location choice model the focus is on the locations, with many routes included. In the model, aspects of the entire route are used because they can help establish why departing pedestrian choose a particular alternative. This means that the entire route of the pedestrian is registered and used: entrance – activity – stairs – platform.

Section 6.1 discusses the research questions relating to the activity location choice model. After that the activity chosen for the location choice model of this study is addressed in Section 6.2. The departing passengers of relevance are selected in Section 6.3. Section 6.4 discusses the factors and processes that will be taken into account in the model. There the definitions of each of the factors and processes are discussed, and a first statistical insight in the relationships is provided. Section 6.5 addresses the results and findings of the activity location choice model. Finally, Section 6.6 presents the conclusions of the model.

6.1. RESEARCH QUESTIONS

The research questions identified for the quantitative research relating to activity location choice are derived from the theoretical framework identified in Chapter 2. The selection on the relationships of interest in Chapter 3 resulted in the following factors:

- Time spent
- Time of day or week
- Distance
- Travel time
- Time table
- Train operations
- Orientation

The time spent can reflect the total time spent in the station building, however in this case the time spent is reflected as the time spent inside the location of the activity. The travel time is divided into the travel time from entrance towards the location of the activity (before), and the travel time from the location of the activity towards the platform (after). Combining the travel time and time spent is not possible, since it is not known what the time spent at the location would be in case where the pedestrian would choose the other location. A visualisation of the time related factors is provided in Figure 33.

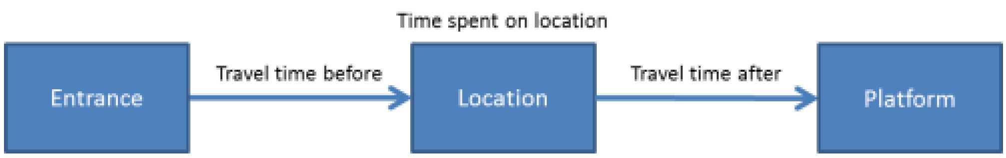

Figure 33: Example on time spent, travel time before and travel time after

The distance is operationalised in two ways. The first is the total distance covered by a pedestrian. The second is whether the pedestrian needs to take a detour to visit the location, relating to the entrance and platform he visits. The time table is also operationalised in two ways: the platform of departure and the service type of the train that was boarded. The train operations are reflected by the delay of a train. The time of day or week is used as a proxy for the familiarity of pedestrians with the train station. This is done by distinguishing between the peak and off-peak hours. As mentioned before, this relationship is not one on one. This means that the proxy does not 100% reflect what familiar and unfamiliar pedestrians do or prefer.

The research questions resulting from the framework, selection and operationalization of the factors are as follows:

1. What is the influence of the *time spent at the location* on the choice for the activity location?
2. What is the influence of the *total distance* on the choice for activity location?
3. What is the influence of a *detour in the route* on the choice of activity location?
4. What is the influence of the *travel time before* on the choice for activity location?
5. What is the influence of the *travel time after* to the platform on the choice for activity location?
6. What is the influence of *peak and off-peak* on the activity location choice?
7. What is the influence of the *service type of the train* that is boarded on the activity location choice?
8. What is the influence of the *platform of departure* on the activity location choice?
9. What is the influence of a *delay of the train* on the activity location choice?
10. What is the influence of the *orientation* on the activity location choice?

6.2. ACTIVITY SELECTED FOR LOCATION CHOICE MODEL

The activity location choice model reflects the choice for a certain location relating to an activity. This means that the activity needs to be selected for the model estimation. The activities that are identified in the station building relating to the shops equipped with SMART station are buying coffee, buying a burger/ fries, buying pasta, getting information/buying a ticket, buying a sandwich, buying something non-food related. These activities are usually possible at multiple locations. Therefore an overview of the locations is provided in Table 29. Some shops have multiple locations, this is indicated between brackets.

Table 29: Locations at which each activity can be performed

Activity	Location 1	Location 2	Location 3	Location 4	Location 5
Coffee	Julia's	Broodzaak	Kiosk (4x)	Starbucks (2x)	AH to Go (2x)
Burger/Fries	Burger King	Smullers			
Pasta	Julia's				
Info/Ticket	Tickets & Service				
Sandwich	Broodzaak	AH to Go (2x)	Julia's	Kiosk (4x)	
Non-food	Rituals	Paperchase			

In order to make a good comparison between locations of the activity, three aspects are important to take into account. The first aspect is that there needs to be a choice amongst locations. This means that the activities buying pasta and getting information or buying a ticket cannot be taken into account in the model. The second aspect relates to the possibility that the activity is performed at that location. If one location serves multiple activities it is not certain that the departing passenger did perform that exact activity, because this is not measured with SMART station. This means that the shops Julia's, Broodzaak, AH to Go and Kiosk are excluded. The third aspect relates to the comparability of the products sold at each location. These products need to be comparable in order to model the activity location choice. In the case of the non-food activity the products sold in Rituals and Paperchase differ largely in their assortment. This means that no proper comparison can be made and that these need to be excluded. These demands result in the exclusion of several activities and locations. The remaining activities and locations are the following:

1. Coffee → Starbucks Bruna, Starbucks BK
2. Burger/Fries → Smullers, Burger King

These activities are compared to each other with a multi criteria analysis (MCA). This defines the best activity for estimating the activity location choice model. The criteria on which the activities are judged are presented in Table 30. Also the demand for the best scoring activity and the weight relating to the criterion are stated. It is expected that distance is important in the model (just as it was for the route choice model), therefore this criteria is provided the highest weight. The distribution compared to SMART station is again important as it reflects the population best. The train related factors are of less importance because the relationship with the location of the activity and the train processes is not as narrow as it was for the route choice model (where the stairs provided direct access to the train).

Table 30: MCA criteria and weight - activities

Criterion	Best score?	Weight
Distance entrance – location - platform	Most variation	5
Distribution of service types	Most variation	3
Delay of trains	Most delay	2
Time spent distribution	Most variation	3
Peak/off-peak distribution versus SMART station	Most comparable	4

Since only two activities are compared in the MCA analysis, the score provided to each activity is 1 or 2. The best activity on each criterion gets the highest score. This score is multiplied with the weight of the criteria to find the total score. The results of the MCA analysis are shown in Table 31. The activity coffee scores best with 4 out 5 criteria and overall is the best scoring activity. For the time spent distribution the activity buying a burger or fries scores better. The activity coffee is selected for the activity location choice model. The analysis for the MCA can be found in Appendix H.

Table 31: Multi criteria analysis on activity selection

Activity	Distance	Distribution service type	Delay	Time spent	Peak/off-peak	TOTAL
Coffee	*10*	*6*	*4*	3	*8*	*31*
Burger/Fries	5	3	2	*6*	4	20

The activity coffee has two locations at which it can be performed (that passed requirements mentioned above): Starbucks Bruna and Starbucks BK. Because the activity location choice model is estimated separate from the route choice model many, routes are possible in the model. Each route leading from an entrance via either of the Starbucks shops to a platform is a route. Two examples of routes are provided in Figure 34 and Figure 35 shows both Starbucks shops.

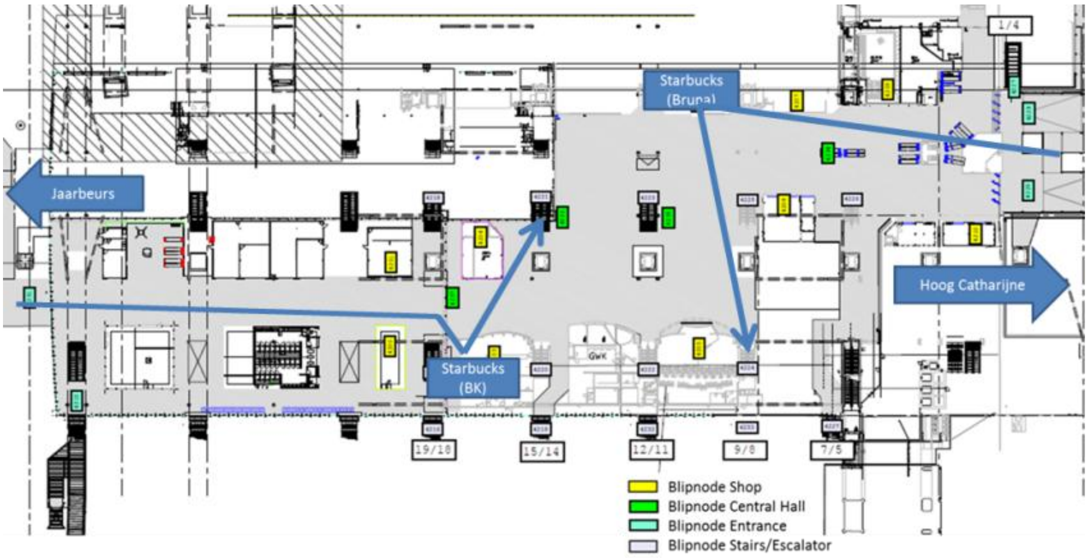

Figure 34: Visualisation of activity location choice

Starbucks Bruna Starbucks BK

Figure 35: Pictures of the Starbucks locations in Utrecht Central station

6.3. SELECTION OF DEPARTING PASSENGERS

Not all departing pedestrians buy a coffee in the station building. Therefore, a selection is needed of the departing pedestrians that are included in the model. This selection process is visualised in Figure 36. The first step of the selection process is to exclude all the pedestrians from the dataset who do not visit the Starbucks shops or that visit both locations. The latter happens in 50 cases. These people were recorded in both shops for more than 30 seconds. This indicates that they either waited in line at one shop and then decided to visit the other shop due to time constraints or they indeed visited both shops and bought a product. After this first selection the dataset is reduced from 240,949 departing pedestrians to 3,893 departing pedestrians.

In order to model the location choice of buying a coffee it is important that the departing pedestrians do not perform other activities. This means that in the second step the other registered activities are filtered. The departing pedestrians that perform other activities are therefore removed from the dataset. After this selection 2,697 departing pedestrians remain in the dataset.

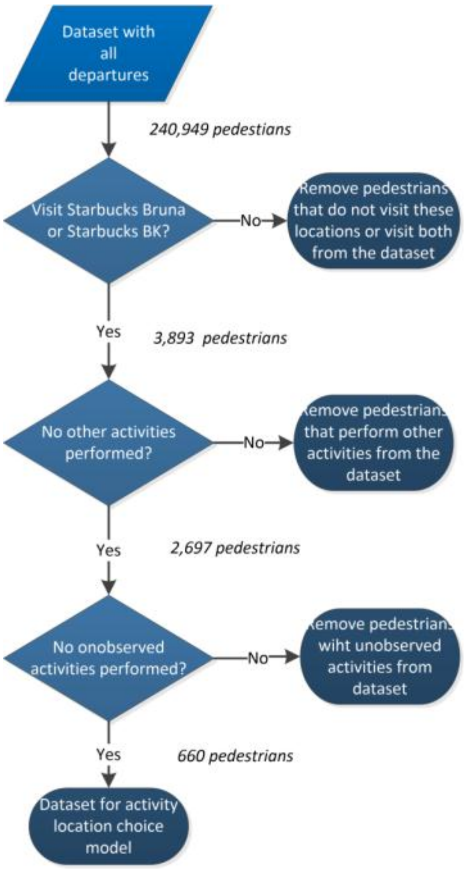

Figure 36: Selection process of departing pedestrians for the activity location choice model

The third step is to remove pedestrians that have performed unobserved activities. As mentioned before, only 55% of the shops are equipped with SMART station. This means that the other 45% reflect shops that people can visit without being registered by SMART station. These unobserved activities are identified by the travel time and distance of pedestrians. The travel time before and travel time after the location visit are considered. Just as in the route choice model a boundary of 3 km/h is introduced (it was determined to be the best limit with respect to type I and type II errors, see Section 5.3). The people that are slower than this from the entrance to the location are removed from the dataset, after this selection 660 departing passengers remain.

When applying the 3 km/h threshold on the travel time from location to stairs and platform an interesting result was found. After this selection only +/- 200 pedestrians remained. This means that most people that visit the Starbucks do not head to the platform straight away. In Utrecht Central station there are benches and wait locations placed in the station building. It is expected that many of the pedestrians that visit Starbucks wait there and finish their coffee. It could also be that these people actually perform other activities, however this is not known from SMART station measurements. The number of departing pedestrians that remain in the dataset is so small that it was decided that all these departing pedestrians were to be included. The total number of pedestrians taken into account in the model estimation is 660. Starbucks Bruna has 533 visitors and Starbucks BK has 127 visitors.

6.4. FACTORS AND PROCESSES INCLUDED IN THE MODEL

In Section 6.1 the factors and processes selected for quantitative research were transformed in operational variables. These variables are the following: travel time from entrance to location, travel time from location to stairs, time spent at the location, peak/off-peak, total distance, detour, service type of the train boarded, platform of departure and orientation. Travel time in general is different for Bluetooth and Wi-Fi users, therefore a check on the differences will be completed. The operationalization of these factors is completed in this section. Also a first analysis on the relationship between these factors and the activity location choice is provided.

TRAVEL TIME FROM ENTRANCE TO LOCATION

The travel time is only measured and registered for the alternative that was chosen. Since the travel time differs per alternative an estimation of the travel time towards the non-chosen alternative is needed. This calculation is based on the (shortest) distance from the entrance to the location and the travel time of a pedestrian over this distance. This way the speed of the pedestrian can be determined, which can be reflected on the other location. The estimation is similar to the time spent calculation for the route choice model. The following formula (when the alternative Starbucks BK is chosen) is applied for the calculation:

$$T_{Br} = \frac{D_{Br}}{\frac{D_{BK}}{T}}$$

Where T_{Br} is the travel time towards Starbucks Bruna, D_{Br} is the distance from the entrance used to Starbucks Bruna, D_{BK} is the distance from the entrance used to Starbucks BK and T is the actual travel time. The distribution of the travel time from entrance to location is shown in Figure 37.

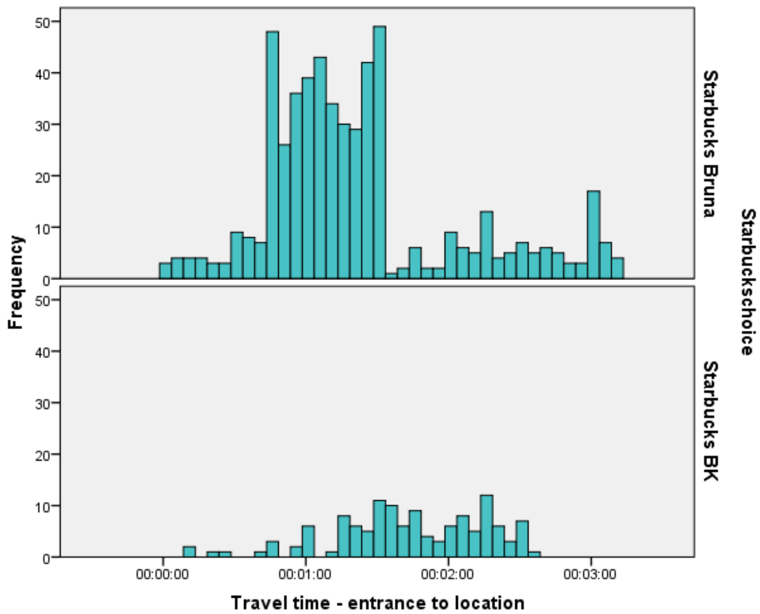

Figure 37: Travel time distribution - entrance to location

It is visible that the most pedestrians that visit Starbucks Bruna have a travel time towards the location that is around 1 minute. The travel time towards Starbucks BK is higher. Most pedestrians travel between 1 and 2:30 minutes. The mean travel time towards Starbucks Bruna is 1:22 minutes with a standard deviation of 0:40 minutes. This mean time for Starbucks BK is 1:43 with a standard deviation of 0:32 minutes. This confirms the difference in travel time identified in the graph. This can be an indication that this factor is of significant influence on the location choice of the departing pedestrian.

Chapter 4 established that the time spent differs for Bluetooth and Wi-Fi users. The difference in amount of users of Wi-Fi versus Bluetooth is rather large (91.8% versus 8.2%). The spreading for both groups is rather large, therefore the median travel time is compared. The median travel time towards Starbucks Bruna is 1:12

minutes for Wi-Fi users and 1:20 minutes for Bluetooth. This means that there is a difference present between the two groups (the mean and variance differ significantly from each other). The median travel time towards Starbucks BK is 1:45 minutes for Wi-Fi users and 1:32 minutes for Bluetooth users. A difference is also present here (however the mean and variance do not differ significantly form each other on a 95% confidence interval). This means that the technology needs to be taken into account in the model estimation.

TRAVEL TIME FROM LOCATION TO PLATFORM

The travel time from the location to the platform cannot be determined for both locations, because both pedestrians heading straight to the platform and pedestrians waiting in the station hall or performing activities afterwards are included. If the same method would be applied as to the travel time from entrance to location, this would provide errors. In case that the travel time is high (a person waited in the station hall for a long time), the travel time from the other location will be significantly higher or lower (depending on the distance) because the estimated speed is very low. This method is not applicable, therefore this factor is taken into account as a generic factor. The distribution of the travel time from location to the platform is presented in Figure 38.

The Starbucks BK has less visitors and the peak amongst pedestrians heading straight to the platform. This is the same for Starbucks Bruna, however more people seem to stay for a longer time period. 35% of the visitors to Starbucks BK head straight to the platform. In case of Starbucks Bruna this is 29%. This means that the difference is rather small. Of the total pedestrians that head straight to the platform 77.9% visit Starbucks Bruna. For the pedestrians waiting in the hall this is 82.0%. A chi-square test on the possible relationship between these variables shows that this is not significant (significance is 0.220 at 95% confidence level). These are therefore independent of each other and do not have to be taken into account in the model estimation.

The mean travel time from location to platform is 6:21 minutes for Starbucks Bruna with a standard deviation of 6:14 minutes. The mean travel time is 5:33 minutes for Starbucks BK with a standard deviation of 6:03 minutes (however no time is registered below 0:30 minutes). This means that differences do indeed exist in the time travelled form location to platform.

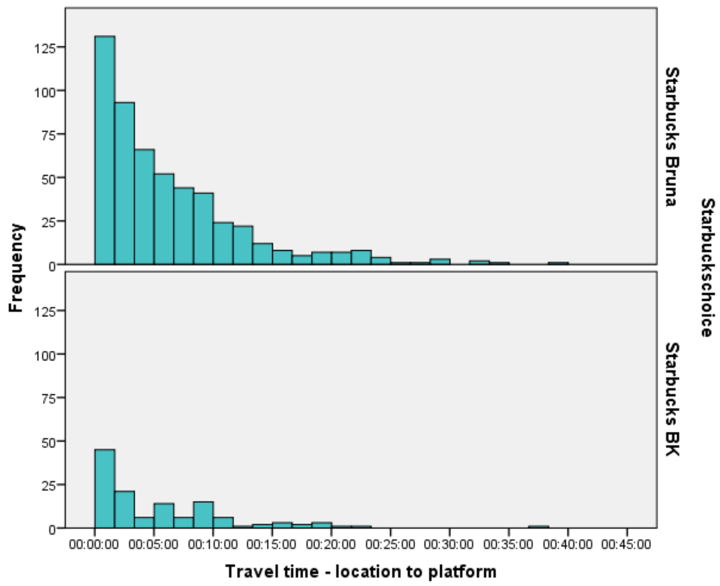

Figure 38: Travel time distribution - location to platform

The same accounts for the differences between Bluetooth and Wi-Fi users. The Bluetooth users have a much lower median time (1:54 minutes for Starbucks Bruna and 1:36 minutes for Starbucks BK) than the Wi-Fi users (4:34 minutes for Starbucks Bruna and 3:51 minutes for Starbucks BK). This means that the difference for Wi-Fi and Bluetooth needs to be taken into account in the model estimation.

TIME SPENT ON THE LOCATION

The time spent on the location is registered for the trip made in the station building. The time spent on the location cannot be determined for the other location, since this might depend on the locations characteristics. Therefore the time spent on the location is taken as a generic attribute. The distribution of the time spent on the location is visualised in Figure 39. The majority of the time spent in Starbucks BK lies between 0:30 minutes and 6:00 minutes. The majority of the time spent in Starbucks Bruna has a wider range, namely form 0:30 minutes to approximately 9:00 minutes.

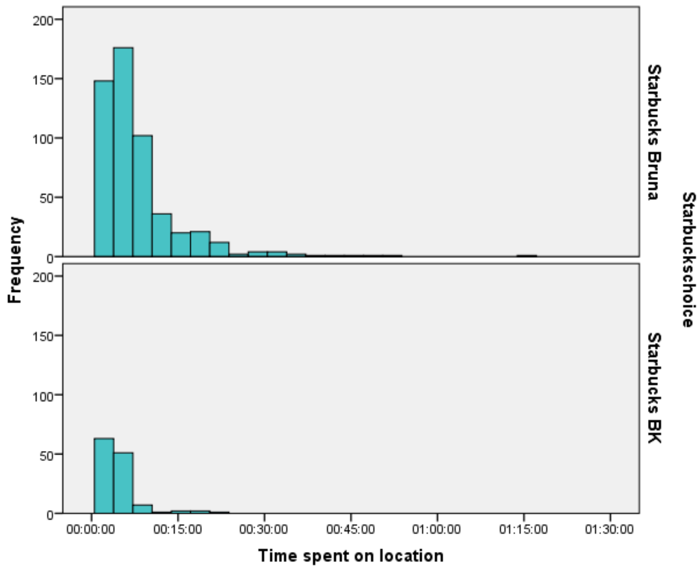

Figure 39: Distribution of time spent on location

The average time spent in Starbucks Bruna is 7:55 minutes with a standard deviation of 7:36 minutes. The mean for Starbucks BK is much lower with 4:26 minutes and a standard deviation of 3:29 minutes. The difference between the time spent for both locations is significant. This factor could be of influence on the activity location choice. The difference between Bluetooth and Wi-Fi users is not significant for the time spent for both locations. This means that the technology does not need to be included in the model estimation regarding the time spent on location.

PEAK/OFF-PEAK

The peak and off-peak hours reflect the time of day or week. The coding of the factor is similar to the route choice model. The distribution peak and off-peak over the location is shown in Table 32. There are differences in the visits of both locations during the day. The Starbucks BK is visited more often during peak hours compared to Starbucks Bruna. This can indicate that this factor indeed influences the location choice.

Table 32: Distribution of peak/off-peak over location choice

Activity location	Peak	Off-peak
Starbucks Bruna	157 (73.7%)	376 (84.1%)
Starbucks BK	56 (26.3%)	71 (15.9%)

TOTAL DISTANCE

The total distance cannot be registered with SMART station. Therefore, the shortest distance needs to be measured. The total distance is the distance from the entrance to the platform via the location and a certain stairs. It is therefore an alternative specific factor. The distances present in the station building are provided in Appendix H (**Fout! Verwijzingsbron niet gevonden.**). The unit for distance used in the model estimation is hectometre. This means that the distribution of distance ranges from 1.05 hectometre up to 2.95 hectometre. This way the factor is in the same range as the effect coded and dummy coded factors. The distribution of departing passengers over these routes shows significant differences. The range is from 0% pedestrians from the same entrance and stairs choosing Starbucks Bruna up to 100%. This factor is therefore expected to have an influence on the choice behaviour.

DETOUR TO VISIT LOCATION

The detour in the visited location can be established when looking at the entrance and platform of departure (including the stairs used) and determining whether the person makes a detour to this location or not. It is therefore possible that a pedestrian needs to make a detour to both locations. The factor is alternative specific as it tells something about the location. The factor is coded using effect coding: no detour is given a value of 1 and a detour is given -1. The distribution of this factor on the locations is shown in Table 33. There is a difference between detours and no detours. In case of a detour, people have a higher preference for Starbucks Bruna than without a detour. This could mean that the factor is of influence on the activity location choice.

Table 33: Distribution of detour over location choice

Activity location	Detour	No detour
Starbucks Bruna	382 (83.2%)	151 (75.1%)
Starbucks BK	77 (16.8%)	50 (24.9%)

SERVICE TYPE OF THE TRAIN

The service type of the train is divided into three groups: intercity, local and international. The factor is generic for both locations since it is related to the train processes. The coding used is effect coding. With three variables this means that two indicator variables need to be used. This is shown below.

	Indicator 1	Indicator 2
IC	1	0
SPR	0	1
INT	-1	-1

The combination of these indicator variables result in a different parameter for each service type. The distribution of these variables over the activity location is shown in Table 34. The distribution differs slightly over the location. However, this result is rather small therefore the impact of the variable is expected to be small.

Table 34: Distribution of service type over location choice

Activity location	INT	SPR	IC
Starbucks Bruna	5 (83.3%)	151 (76.6%)	377 (82.5%)
Starbucks BK	1 (16.7%)	46 (23.4%)	80 (17.5%)

PLATFORM OF DEPARTURE

In Utrecht Central station 6 platforms are present. The platform of departure reflects the train processes and is therefore generic. This factor is also coded using effect coding. With 6 platforms, this means that 5 indicator variables are introduced (k-1). This is shown below.

	Indicator 1	Indicator 2	Indicator 3	Indicator 4	Indicator 5
1-4	1	0	0	0	0
5/7	0	1	0	0	0
8/9	0	0	1	0	0
11/12	0	0	0	1	0
14/15	0	0	0	0	1
18/19	-1	-1	-1	-1	-1

The distribution of this factor over the locations is shown in Table 35. The preference towards Starbucks Bruna is stronger for platforms 1-4, 5/7 and 8/9. Platform 11/12 and 14/15 are located close to both, therefore the preference is less extreme for those two. For platform 18/19 this preference is even lower.

Table 35: Distribution of platform over location choice

Activity location	1-4	5/7	8/9	11/12	14/15	18/19
Starbucks Bruna	43 (87.8%)	115 (87.1%)	88(91.7%)	111 (83.5%)	103 (78.6%)	73 (61.3%)
Starbucks BK	6 (12.2%)	17 (12.9%)	8 (8.3%)	22 (16.5%)	28 (21.4%)	46 (38.7%)

DELAY OF THE TRAIN

The train is delayed when it departs more than 3 minutes later than scheduled. The coding is similar to route choice model. The distribution of delayed and on time trains is provided in Table 36. The differences between boarding a train that is on time versus a train that is delayed is very small. Therefore the impact of this factor is expected to be small.

Table 36: Distribution of delay over location choice

Activity location	Delayed	On time
Starbucks Bruna	88 (86.3%)	445 (79.9%)
Starbucks BK	14 (13.7%)	113 (20.1%)

ORIENTATION

The orientation reflects the preference for left or right relating to the location of the activity (as seen from the entrance). This factor is again coded using effect coding: right is 1 and left is -1. The distribution is shown in Table 37.

Table 37: Distribution of orientation over location choice

Activity location	Left	Right
Starbucks Bruna	119 (62.3%)	414 (88.3%)
Starbucks BK	72 (37.7%)	55 (11.7%)

The preferences do differ largely between left and right. People choose Starbucks BK more often when it is located left, whereas Starbucks Bruna is more often chosen when located right. Therefore, impact is expected on the location choice.

In summary, based on the first analysis several factors are expected to have a large influence on the activity location choice. These are travel time before, travel time after, time spent on location, peak/off-peak, distance, detour and orientation. It is expected that the service type, platform and delay are of less influence.

6.5. RESULTS AND FINDINGS

This section addresses the activity location choice model. This is estimated using the software package BIOGEME (Bierlaire, 2003). The model estimation is completed in several steps, just as for the route choice model. In this case the multinomial logit model is again estimated as the base model, since it is the most simple. The best multinomial logit model can possibly be improved by introducing a mixed logit model. The multinomial logit model is estimated using the following steps:

1. Estimate a model for each of the factors and processes individually
2. Combine the significant factors and processes in a stepwise manner to create the best model

The first step is to estimate a model for each factor and process individually. This provides a good overview of the factors and processes that have significant influence on the activity location choice and which can be combined into the best combined model. In each model an ASC is also estimated, which shows the base preference of the departing pedestrian towards the locations and includes unobserved attributes. Appendix I shows an extensive overview of all the model results. Table 38 shows an overview of the individual models estimated.

It was identified that Wi-Fi and Bluetooth users have a different travel time before and after, therefore this factor should also be taken into account. However, in order to calculate the interaction effect present for this factor, the technology must also have an influence on the location choice. Therefore, the technology is estimated as an individual parameter first. The technology does not influence the location choice on its own and therefore the interaction cannot be estimated (no interpretation possible).

Table 38: Overview of individual models estimated for activity location choice – parameters

Parameter	Value	Robust t-test	p-value	Log-likelihood	Adjusted rho-square
Travel time before	-0.408	-3.62	0.00*	-316.489	0.304
Travel time after	$3.33e^{-15}$	0.00	1.00	-323.215	0.289
Time spent on location	$3.88e^{-15}$	0.00	1.00	-323.215	0.289
Peak	$1.73e^{-16}$	0.00	1.00	-323.215	0.287
Off-peak	$3.63e^{-16}$	0.00	1.00		
Distance	-1.21	-7.70	0.00*	-288.412	0.365
Detour	-0.424	-5.62	0.00*	-306.313	0.326
Service type 1	$3.66e^{-16}$	0.00	1.00	-323.215	0.287
Service type 2	$1.55e^{-16}$	0.00	1.00		
Platform 1	$-5.68e^{-17}$	0.00	1.00	-323.215	0.280
Platform 2	$1.06e^{-17}$	0.00	1.00		
Platform 3	$-1.87e^{-17}$	0.00	1.00		
Platform 4	$1.14e^{-17}$	0.00	1.00		
Platform 5	$9.74e^{-18}$	0.00	1.00		
Delay	$-3.70e^{-16}$	0.00	1.00	-323.215	0.289
Orientation	0.244	4.72	0.00*	-312.426	0.313
Technology	$4.48e^{-16}$	0.00	1.00	-323.215	0.289

*Significant on a 95% confidence interval

Six of the estimated factors do not have a significant influence on the activity location choice. These factors are travel time after, time spent on location, peak/off-peak, service type, platform and delay. The value that indicates the effect on utility does not differ significantly from zero. This also means that six of the research questions can already be answered.

The travel time after does not influence the location of the activity. Introducing an alternative specific parameter did not provide any better results. This factor differs for both locations and is now taken into account as a generic factor. This was done because no estimation of the travel time after was possible for the other location. However, since there are differences related to location choice it is expected that the factor is alternative specific. This can however not be tested in this model using this data. The same applies to the time spent on location. An alternative specific parameter did not improve the model. It is possible however that these factors are reflected in the ASC, as they cannot be captured in the model but do provide differences. The peak and off-peak are of no significant influence on the location choice. Introducing alternative specific parameters for this factor did make the parameters significant, however the model was unidentifiable. Therefore, this provided no improvement to the model. The peak and off-peak reflect a proxy for the familiarity of the departing passenger. It now seems that for this proxy the familiarity of pedestrians does not influence their location choice. But, as this relationship is not one on one, it cannot be concluded that the familiarity of the pedestrian does not influence activity location choice. The service type does not influence the location choice, which means that the pedestrians do not choose different if they board a different train service. The platform can be seen in the same way. Regardless of the platform they go to they make their decision on the location. The last factor is the delay of the train.

The relationships identified and rejected showed an interesting pattern. All factors that influence the location before it is visited are of significant influence, whereas the factors of influence after the location is visited are not. This seems to indicate that departing pedestrians make their decisions based on the part of the route before the location is visited and not based on the part of the route after the location is visited.

Four factors are estimated to have a significant influence on the activity location choice (95% confidence interval): travel time before, total distance, detour and orientation. The travel time before parameter has a value of -0.408, which means that with one extra minute 0.408 utility points are subtracted. The distance parameter has a value of -1.21. The effect of this factor on the location choice is strongest. If a pedestrian needs to make a detour to visit the location this reduces the utility of that location with 0.424. As the Dutch population is generally oriented to the right it is expected that the location is valued higher on the right side.

As mentioned before, the model assumes that all the pedestrians know of both locations and consider them both. However, some pedestrians might not be aware of both locations as they might be unfamiliar or do not have to cross the entire station and therefore not notice both alternatives. This needs to be taken into account when making conclusions.

The second step is to combine all significant factors into the best model. The base for this model is the distance, as it has the highest added value. The distance is the most dominant factor in the model. This results in the fact that no better model can be estimated for the activity location choice, based on the factor distance. When adding the factor detour to the model, it becomes not significant (p-value of 0.47). This same happens to the factor orientation (p-value of 0.07). When adding the travel time to the model it is significant and provides a better model, however there is a high correlation between the parameter distance and the parameter travel time. In addition, due to the dominance of the factor distance, the sign of the travel time variable is positive. This means that mathematically the model will be correct, but it cannot be read properly. This basically means that the best model only includes distance. Introducing alternative specific parameters does not improve the model. Therefore the utility function becomes:

$$U_i = ASC_i + \beta_d \times Distance_i$$

The general information of this model is shown in Table 39. The likelihood ratio test of this model is 338.130 which is higher than the threshold of 5.99 at 95% confidence interval. This means that the model is better than the null situation. In addition, the model has an adjusted rho-square of 0.365. When comparing this value to the adjusted rho-square of the models with only the ASC as indicator (0.289), the model is better. The adjusted rho-square increases with 0.076.

Table 39: General information on the best MNL model

Model	Multinomial Logit
Number of estimated parameters	2
Number of observations	660
Number of individuals	660
Null log-likelihood	-457.477
Final log-likelihood	-288.412
Likelihood ratio test	338.130
Adjusted rho-square	0.365
Diagnostic	Convergence reached...
Iterations	10

The utility parameters of this model are provided in Table 40. The ASC of Starbucks Bruna is fixed to zero. The ASC of Starbucks BK indicates that this location is less preferable by the pedestrians. This is true when looking at the distribution of the travellers over the locations. The parameter for distance is -1.21 and the range of the variable distance is from 1.05 to 2.95. This means that the impact of the attribute ranges from -1.27 to -3.56. This means that the importance of distance is rather large (compared to the ASC).

Table 40: Utility parameters of the best MNL model

Parameter	Value	Robust std error	Robust t-test	p-value
ASC_BK	-1.20	0.105	-11.40	0.00
ASC_BR	0.00			
Beta_d	-1.21	0.157	-7.70	0.00

An example is provided below to illustrate the working of the model. The example shows the importance of the factor distance once more.

Pedestrian B comes from Hoog Catharijne. He buys a coffee at the Starbucks located close to Bruna. After this he heads towards platform 8/9 and takes the Mid stairs. The following factor values are found for this pedestrian:

- Distance via Starbucks Bruna = 1.30 hectometre
- Distance via Starbucks BK = 2.10 hectometre

The impact of the factor distance on the total utility of Pedestrian B is the following:
$U_{\text{Starbucks Bruna}}$ = -1,573
$U_{\text{Starbucks BK}}$ = -1,20 – 2.541 = -3.741

Pedestrian B has a probability of 89.7% for choosing Starbucks Bruna and a probability of 10.3% of choosing Starbucks BK. He chooses Starbcuks Bruna.

The remaining four research questions can now also be answered. The travel time before is only estimated in the individual model. The factor has a range of 0.02 to 3.18 minutes. The minimum travel time is therefore equal to 3 seconds. This is missed by the filter (45 seconds) because the filter on time spent was applied to the total time spent in the station hall. In this case the pedestrian has a higher total time spent than 45 seconds. It could therefore be that he was first scanned while he was already inside the building. This would explain why 3 seconds are measured. However, this needs to be taken into account as the accuracy of the Wi-Fi scanners is not that high and these low measurements are done by Wi-Fi scanners. The range of impact is then -0.008 to -1.297 (average is -0.472). Given the ASC for Starbucks BK (-1.35) this impact ranges from very little to rather high.

The distance provides the most dominant impact on the total utility with a range from -1.27 up to -3.56. The detour present in the model subtracts 0.424 of the total utility, whereas a direct line provides 0.424 utility. The impact of this factor is therefore much less than the distance (3 to 8 times less). Regarding the travel time this factor provides 3 times less utility in the maximum situation, but 53 times more in the minimum situation. The orientation provides 0.244 to the total utility if the location is on the right side and subtracts 0.244 in case of a location on the left side. The influence of this factor is lowest. The orientation is therefore less important than the travel time towards the shop, whether it is a detour and the total distance.

The fact that the total distance is considered that important means that pedestrians are actually optimising their whole route within the station building. This is also the result when looking at the detour. This seems to indicate that the departing pedestrian can choose the location and simultaneously optimise the route. Which is as expected. However, this conclusion cannot be made without investigating the combined process.

As the group was selected on their matching activities it is expected that the preferences of the group are homogeneous related to the total distance. The mixed logit model estimated indeed confirms this. This means that the individual MNL model for distance is the best estimated model.

6.6. CONCLUSIONS

Activity location choice is the choice for a certain location to perform an activity. In order to identify the factors and processes of influence on this choice, all departing pedestrians that perform that activity on either of the locations are taken into account. This means that people with multiple origins and destinations are included. The location where the activity is performed is modelled separate from the route choice, in order to establish what determines the location chosen. However, in determining the location choice the entire route is registered and used: entrance – location – stairs – platform. The activity selected for the location choice model is buying a coffee. By means of a multi criteria analysis it was determined that this activity was best suitable. The locations included in the model are Starbucks Bruna and Starbucks BK. The other locations that sell coffee facilitate multiple activities. Therefore, it cannot be established whether a coffee or another product is bought. These locations are therefore excluded from the model.

In Chapter 3 factors and processes are selected that possibly influence activity location choice behaviour of departing pedestrians. These relationships are quantified by means of a discrete choice model. The factors concerned are time spent, time of day or week, travel time, distance, time table, train operations and orientation. The time spent is used as the time spent on the location. Travel time is separated into two parts, travel time from entrance to location and travel time from location to platform. The time of day or week is reflected as travelling during peak hours or during off-peak hours (just as in the rout choice model). The distance is reflected by two factors: the total distance (entrance – location – stairs) and whether the pedestrian needs to make a detour to visit the location. The time table is reflected by two factors: the service type of the train boarded and the platform of departure. Finally, the train operations are reflected by the delay of the train boarded.

Based on a first analysis on the relationship between the factors and processes and the coffee location choice, a significant influence is expected from travel time before, travel time after, time spent on location, peak/off-peak, total distance, detour and orientation. When estimating the model the travel time before (no difference were found in the model estimation for Wi-Fi and Bluetooth users), total distance, detour and orientation are of significant influence. The travel time after and time spent on location do differ significantly for both locations, however it was not possible to estimate these times for the non-chosen alternative. This means that they are included in the model as generic attributes. The influence from the factor on the location choice was then not significantly different from zero. It could be that these factors are part of the ASC of the model, as they do differ. The peak and off-peak also turn out not to influence the location choice. Regarding this proxy the familiarity of the pedestrians does not matter for choosing a location. The relationship between the proxy and the familiarity however is not one on one, therefore it cannot be concluded that the familiarity of the pedestrians has no influence on the location choice. Related to this aspect is the considered choice set. Unfamiliar pedestrians might not include both alternatives in the choice set and therefore make a choice in a different way. This cannot be captured by the model.

Combining all significant the factors into one model however is not possible. The distance is so dominant that the orientation and detour can be neglected when combined with this factor. The travel time is still significant when combined with the distance, however the model becomes unreadable. Therefore the total distance predicts the activity location choice in the best way.

Since the factors cannot be combined in one model the direct relationship between the factors cannot be derived. However, the distance has an impact that is 3 to 8 times higher than the detour. The impact of the orientation on the location choice is lowest (5 to 14 times lower than distance). The preference goes out the right side, the location is more likely to be chosen when located right (from the entrance). The travel time before shows a lowest value of 3 seconds. The limit in the data filtering is set to 45 seconds, the total time spent for this pedestrian is higher than this value. Therefore the pedestrian was not removed from the dataset. 55 persons have a travel time towards the location that is lower than 45 seconds. This means that for some of these persons the measured time is not accurate (Wi-Fi inaccuracy). This means that this needs to be taken into account when deriving conclusions. The impact from the travel time is lower than the impact from distance (3 to 53 times). The 53 times arise from the very low travel time, which is not accurate.

The total distance has the largest influence on the location choice of the activity. The other factors that are of significant influence are also aimed at either the entire route through the station building or the part of the route before the visit. The insignificant factors are aimed at after the location is visited. This seems to indicate that the pedestrian optimises the entire route specifically aimed at the first part of the route. A detour in the visit is disliked, the travel time to the alternative is optimised and the total distance is reduced. This statement is confirmed by the insignificant factors. The platform itself does not influence the location choice, the combination of entrance and platform (reflected by distance) does. The service type of the train that is boarded also does not influence the location choice. It does not seem to matter which type is boarded (intercity, sprinter or international train). The delay also does not influence the location choice.


11b
12b

7. CONCLUSIONS AND RECOMMENDATIONS

In this chapter conclusions are derived from the research (Section 7.1). This is done by providing an answer to the research question. After that the discussion is addressed in Section 7.2. In this section the findings and issues of this research are discussed. Finally, recommendations are made for science (Section 7.3) and practice (Section 7.4).

7.1. CONCLUSIONS

This research aims to explain and predict how departing pedestrians in the train station choose their route and activity location. The main research question identified in this research is the following:

'Which factors and processes influence the route and activity choice behaviour of pedestrians in train stations in the process of entering the station building to boarding the train and to which extent?'

From literature and observations in the train station many factors and processes were identified that influence or are expected to influence the choice behaviour of departing pedestrians. The factors and processes were assigned to several categories: personal, system, public transport or external. The personal factors are divided into three sub categories: general personal, trip factors and learning processes. The system factors are split up in location factors and route attributes. The familiarity of a pedestrian, belonging to the general personal factors, was addressed in many studies and was found to be of significant influence on the choice behaviour of pedestrians. This means that a familiar pedestrian makes choices in a different way than an unfamiliar person.

Due to the large amount of relationships identified, a selection was needed for the quantitative analysis. The relationships were tested on two criteria. The first is the availability of knowledge. If no (quantitative) research has been found in literature, there was no knowledge available. These relationships then provide opportunities for this study. The second is the ability of SMART station to capture the relation. SMART station is the data collection method used in this study. It tracks and counts pedestrians that are present in the station. The counting is done by means of infrared scanners at the entrances. The tracking is based on Bluetooth and Wi-Fi. Several scanners are located in the station building which register the presence of devices with Bluetooth or Wi-Fi enabled (a discussion on the suitability of SMART station data can be found in Section 7.2). The relations that are selected for the quantitative study are shown in Table 41. The cross in the table indicates the potential influence of the factor on the choice.

Table 41: Factors tested in the quantitative research

Factor	Route choice	Activity location choice
Time spent	X	X
Travel time	X	X
Time of day or week	X	X
Distance	X	X
Time table	X	X
Train operations	X	X
Orientation	X	X
Visibility	X	

The relationships are tested by means of data collected at Utrecht Central station. The method used to quantify the relations is discrete choice modelling based on the concept of utility maximisation. This method can be used to estimate if a factor has a significant influence on the choice and how strong the influence is of each factor (also compared to the other factors).

Two models are estimated. The route choice model takes pedestrians into account that head straight to the platform. The activity location choice model takes the pedestrians into account that first perform that activity.

ROUTE CHOICE

The factors shown in Table 41 are not all operational. Therefore they need to be translated to operational factors. In the route choice model pedestrians heading straight to the platform are included. This means that time spent and travel time are the same. The time of day or week is translated to peak and off-peak hours. This factor serves as a proxy for the familiarity of the pedestrian. Pedestrians travelling during peak hours are usually familiar, whereas pedestrians travelling during off-peak hours are usually unfamiliar. The time table is reflected by the stop location of the train. At Utrecht Central station trains can stop on two phases (A and B). A route does not provide access to both phases. Therefore this factor shows whether a route provides direct access to the train or not. The train operations are reflected by the delay of the train.

The operationalised factors are included in the model estimation and the results are visualised in Figure 40. The travel time is the most dominant factor concerning route choice. The longer the travel time, the lower the probability that the route is chosen. This finding supports many of the theories and models estimated for pedestrian route choice behaviour (e.g. Hoogendoorn & Bovy (2002)). The travel time is highly correlated to the shortest distance. Distance therefore also provides a significant influence regarding route choice. The influence of this factor was lower than of travel time. This might be due to the fact that only the shortest distance could be used and not the actual distance. This however also means that these two factors cannot be estimated in the same model. The orientation and train stop location also provide significant influence to the route choice. If the route provides direct access to the train this means that the probability for choosing that route is higher. The influence of this factor compared to the travel time is limited. In addition, the orientation of the route provides significant influence on the route choice. If aimed/located on the right the probability of choosing that route increases. Again the influence of this factor is limited compared to travel time. In conclusion, when choosing a route the departing pedestrian bases his choice mainly on the travel time (and distance), but also takes the orientation and stop location of the train into account.

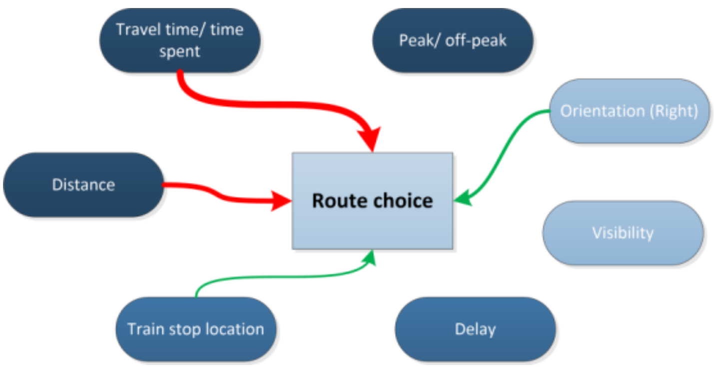

Figure 40: Factors influencing route choice

The route choice does, on the other hand, not depend on the delay, visibility and peak versus off-peak hours. If a train is delayed or not does not influence the route the pedestrian takes in the station building. The peak and off-peak hours do not influence the route choice. In peak hours more pedestrians are present in the station building, however this does not seem to effect the route choice. This factor was used as a proxy for the familiarity of the pedestrians. However, as the relation is not one on one, it cannot be concluded that there is no difference in route choice regarding familiarity. Finally, the visibility of the route does not seem to influence route choice.

ACTIVITY LOCATION CHOICE

The factors in Table 41 need to be translated to operational factors. The travel time is split up in two parts: the travel time from the entrance to the location and the travel time from the location to the platform. The time spent is considered for the time spent on the location. The distance is translated into the total distance from entrance via location to platform and whether a detour is needed for visiting the location. The time table is reflected by the platform of departure and the service type of the train (intercity, sprinter or international). The train operations are again translated into the delay.

The operationalised factors are included in the model estimation and the results are shown in Figure 41. The most dominant factor that influences the activity location choice is the total distance. The larger the total distance, the lower the probability of choosing that location. This indicates that the departing pedestrian optimises his entire route in choosing the location of his activity. The fact that the travel time from entrance to the location also has significant influence on this choice seems to indicate that the optimisation is mainly aimed at the first part of the route. This is confirmed by the fact that the orientation is of significant influence. If the location is present on the right side of the pedestrian as seen from the entrance, the probability for choosing that location increases. A detour present in the total trip from entrance to platform therefore also indicates a lower probability for choosing that location. Concluding it can be said that the pedestrian makes his choice of location taking the entire route into account and specifically optimising the first part of the route.

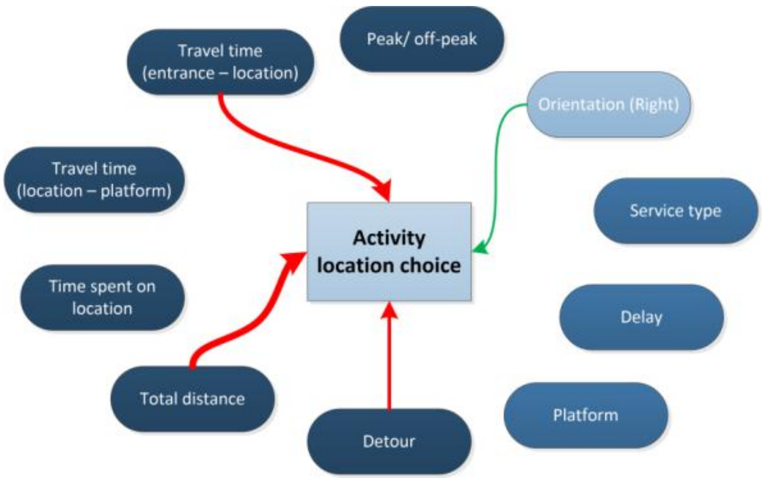

Figure 41: Factors influencing activity location choice

This finding is supported by looking at the factors that are of no significant influence the activity location choice. These factors are the travel time from location to platform, the time spent on location, the platform of departure, the delay, the service type of the train and travelling during peak hours or off-peak hours. They are mainly aimed on the second part of the route. Regarding the peak and off-peak hours, which serve as a proxy for the familiarity of the departing pedestrian, it can be said that this is of no influence on the activity location choice. However, it cannot be concluded that the familiarity does not influence this choice, as this relation is not one on one.

7.2. DISCUSSION

In this section a discussion of the research is addressed. Several points are discussed: the suitability of SMART station data for this research, the generalisation of the results within Utrecht Central station, the generalisation of the results to other stations, modelling familiarity, the considered choice set and two separate models versus one extensive model.

SUITABILITY OF THE SMART STATION DATA FOR THIS RESEARCH

The data collection method SMART station has some limitations regarding the data it collects. Due to privacy issues personal data are not registered. This means that relationships concerning socio-economic data cannot be captured by the data. In addition, the data are of a binary nature. This means that the method only registers if a pedestrian is present at the location or not. The result of this limitation is that the exact movements of the pedestrian cannot be registered. This means that relationships concerning the exact route of the pedestrian cannot be tested. This results in the fact that not all the relationships, of which no knowledge is available, can be captured by means of SMART station data. Section 3.2.3 showed that approximately half of the interesting relations could be captured. This limits the suitability of the method. As this study was faced with a time constraint, not all factors and processes could have been estimated. This however resulted in estimating the factors of influence regarding the trip, system and public transport.

GENERALISATION OF THE RESULTS WITHIN UTRECHT CENTRAL STATION

The route choice model is estimated for pedestrians that head straight to the platform. The model was only estimated for one platform (11/12). It was established in Chapter 4 that the platforms are rather different from one another. Regarding the train processes they can be divided into three groups: final/start platform, intercity platform and mixed platform. Platform 11/12 is an intercity platform. This means that regarding the train processes there might be different results found at other platforms. However, it is expected that in those cases the direct access from the route to the train is also of influence on the route chosen. The platforms are also not all equal in the number of stairs and escalators providing access to them. However it is expected that the orientation, time spent/travel time and distance are also of influence. Even though differences arise in the routes and platforms within Utrecht Central station, it is expected that the same relationships will be found on those.

The activity location choice model is estimated for people buying a coffee. The locations taken into account are Starbucks Bruna and Starbucks BK. The other locations that sell coffee are excluded from the model estimation as they also sell other products. This means that it cannot be established whether the departing pedestrian bought coffee or something else. However, concerning other activities and even other location for buying coffee, it is expected that the orientation, total distance, detour and travel time from entrance to location are of influence. The effect of the total distance and travel time are expected to be less dominant if the locations of the activities are geographically close to each other.

GENERALISATION OF THE RESULTS TO OTHER TRAIN STATIONS

The relationships found in the models on route choice and activity location choice are probably also applicable on other train stations. The lay-out of the differ largely. In case of Utrecht Central station the platforms are entered via one main station building. This means that weight of the factors and processes on the choice will differ from Utrecht. It is however expected that the same findings can be done on other stations.

MODELLING FAMILIARITY

The familiarity is modelled by means of a proxy: peak hours versus off-peak hours. This proxy can never serve as a complete view of the familiarity. The familiar pedestrian also travels during off-peak hours and unfamiliar pedestrian travels during peak hours. In general, it will be as the proxy indicates. However, it would be better to find a factor that captures the familiarity entirely. The conclusions derived for the proxy therefore cannot be translated to conclusions on the familiarity.

CONSIDERED CHOICE SET

The considered choice set regarding route choice or activity location choice is not necessarily the same for all departing pedestrians. Especially regarding pedestrians that are unfamiliar with the station building. They might more often choose for an alternative that is close by than for the alternative that is not visible. In the models estimated it was assumed that every departing pedestrian was aware of the alternatives. The discrete choice models cannot handle the fact that not everybody is aware of the alternatives. Therefore, it is important to create a choice set that is considered by all the pedestrians included. Regarding route choice it is expected that the unfamiliar pedestrians might be unaware of the non-visible routes. This means that the choice set might not be the same for those pedestrians.

TWO SEPARATE MODELS VERSUS ONE EXTENSIVE MODEL

Two separate models were estimated on route choice and activity location choice. Relating to the familiarity of pedestrians there is a difference between the decision making of these pedestrians: simultaneous versus sequential. The familiar pedestrian is expected to make a simultaneous decision. In that case it would be better to combine the two models into one extensive model. However, in the case where no activities are performed by the pedestrian it is possible to estimate a separate model. This also accounts for the activity location choice when including all the routes visiting the location. When including all these routes as alternatives in the model, the sample of pedestrians performing that route will not be high enough. Therefore, combining them into one activity location choice model seems plausible.

7.3. RECOMMENDATIONS FOR SCIENCE

In this section recommendations for future research are provided. These are based on issues that arose during this study.

- The theoretical framework identified many choices that pedestrians need to make in the station building. The factors and processes of influence on these choices provided input for many relationships, of which most are not yet quantified. This means that multiple opportunities for future research arise, for example:
    - Decision making of departing pedestrians before they start their trip, related to the factors and processes of influence.
    - A quantification of the relationships between the choices before the trip.
    - Researching to what extent departing pedestrians plan their trip in advance.
    - Researching how departing pedestrians update/adapt their activity sequence during the trip.

- This study estimates two separate models on the route choice and activity location choice behaviour of departing pedestrians. It would be interesting to combine this into a super network model where both choices are included. From the activity location choice model could be derived that the pedestrians optimise the whole route. Therefore it is expected that a joint model could also be estimated. A comparison between these two approaches could be made to see which model provides better results.

- The discrete choice models were estimated using the concept of utility maximisation. This concept assumes that the decision maker chooses the alternative that provides the highest utility. However, it is not known whether people actually make decisions that way. Other concepts have been developed that approach the discrete choice models in a different way. One example is the regret minimisation concept (introduced by e.g. Chorus et al. (2008)). This concept assumes that the alternative is chosen that leads to the least regret. Transport is usually considered as a mandatory activity which costs time and money. Therefore, regret minimisation seems plausible for transport related researches. It is interesting to compare the two concepts on route and activity location choice models in the station area.

- Discrete choice models based on the concept of utility maximisation assume that the decision maker is rational concerning the choices he makes. This however is not always the case, as optimising the choice costs a lot of effort. Therefore a satisfactory alternative could also be chosen. Research by Zhu (2008) related to the bounded rationality of pedestrians. Bounded rationality relates to the fact that not all alternatives are known by people or considered by them. Even if they are considered, pedestrians are bounded in their ability to choose the best alternative. Therefore, this methodology seems promising.

- The data collected by SMART station is revealed preference. This means that the data shows what has been done by the pedestrians. This means that only data during the trip are collected. It would be interesting to combine these data in some way with stated preference data on the planning of the trip or the feedback relation (how is the process of today incorporated for the next visit). Due to privacy issues, it is not possible to make a one on one relation. However, it could be interesting to match these data to create a broader understanding of the pedestrian choice behaviour.

7.4. RECOMMENDATIONS FOR PRACTICE

This section provides recommendations for practice. These are based on the findings done in this study.

- Departing pedestrians base their route choice mainly on the travel time/distance. This means that regarding the distributions of departing pedestrians over the stairs, it cannot be assumed that they will distribute equally over all alternatives. Also the location of the stop location of the train is of influence on this distribution, meaning that this could be used to optimize the distribution over the stairs.

- Departing pedestrians mainly base their choice for the activity location on the total route within the station building. Therefore the location of the shops in the station can be optimised. When multiple locations are present in the station building that sell a certain product, the total distance is relevant. By locating a shop in the most busy flow (entrance – platform), the location becomes more attractive for visitors. Also, the right side, as seen from the entrance, is preferred.

- The data collected by means of SMART station are generally representative for the population of departing pedestrians in Utrecht Central station. This means that the dataset is valuable for other research relating to departing pedestrians as well.

- SMART station is now installed on multiple stations in the Netherlands. This means that these datasets can be linked. This provides more opportunities regarding, for example, the matching of the data to train process data (if seen at the other station, it is certain that pedestrians boarded a train from platform x at Utrecht Central station) or tracking the total visit of a pedestrian.

REFERENCES

Antonini, G., Bierlaire, M., & Weber, M. (2006). Discrete choice models of pedestrian walking behavior. *Transportation Research Part B*, 667-687.

Bakker, P., & Zwaneveld, P. (2009). *Het belang van openbaar vervoer: De maatschappelijke effecten op een rij.* Den Haag: Centraal Planbureau en Kennisinstituut voor Mobiliteitsbeleid.

Ben-Akiva, M., & Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. In R. W. Hall, *Handbook of transportation science* (pp. 5-33). New York: Springer US.

Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and Application to Travel Demand.* Camebridge : The MIT Press.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. *Proceedings of the 3rd Swiss Transportation Research Conference.* Ascona: Switzerland.

Bierlaire, M. (2009). *Estimation of discrete choice models with BIOGEME 1.8.* Lausanne: Ecole Polytechnique Federale de Lausanne.

Bierlaire, M., & Robin, T. (2009). Pedestrian choices. In H. Timmermans, *Pedestrian behavior: Models, data collection and applications* (pp. 1-26). Bingley, United Kingdom: Emerald.

Bishop, I. D., Wherrett, J. R., & Miller, D. R. (2001). Assessment of path choices on a country walk using a virtual environment. *Landscape and Urban Planning*, 225-237.

Bogers, E. (2009). *Traffic infromation and learning in day-to-day route choice.* Delft: TRAIL.

Borgers, A., & Timmermans, H. (1986). A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas. *Geographical Analysis*, 115-128.

Bovy, P., & Stern, E. (1990). *Route choice: Wayfinding in transport networks.* Kluwer Academic Publishers: Dordrecht.

Cheung, C. Y., & Lam, W. H. (1998). Pedestrian route choices between escalator and stairway in MTR stations. *Journal of Transportation Engineering*, 277-285.

Chorus, C. G., Arentze, T. A., & Timmermans, H. J. (2008). A random regret-minimization model of travel choice. *Transportation Research Part B*, 1-18.

CU2030. (2013). *OV-terminal Utrecht Centraal*. Retrieved December 5, 2013, from CU 'K zie je in Utrecht: http://www.cu2030.nl/page/ovterminal

Daamen, W. (2004). *Modelling passenger flows in public transport facilities.* Delft: Delft University Press.

Daamen, W., & Hoogendoorn, S. P. (2003). Research on pedestrian traffic flow in the Netherlands. *Proceedings Walk 21 IV* (pp. 101-117). Portland: Walk 21 conference.

de Vos, S. (2011). *Transfer Concept.* Utrecht: NS Poort.

Dijkstra, J., Timmermans, H., & de Vries, B. (2009). Modeling impulse and non-impulse store choice processes in a multi-agent simulation of pedestrian activity in shopping environments. In H. Timermans, *Pedestrian behavior: Models, data collection and applications* (pp. 63-86). Bingley, United Kingdom: Emerald.

Ettema, D., Borgers, A., & Timmermans, H. (1993). Simulation model of activity scheduling behavior. *Transportation Research Record*, 1-11.

Fruin, J. J. (1971). *Pedestrian planning and design.* Michigan: Metropolitan Association of Urban Designers and Environmental Planners.

Golledge, R. G. (1999). *Wayfinding behavior: Cognitive mapping and other spatial processes.* Baltimore: Johns Hopkins University Press.

Helbing, D. (1997). *Verkehrsdynamik.* Berlin Heidelberg: Springer - Verlag.

Helbing, D., Farkas, I., Molnar, P., & Vicsek, T. (2002). Simulation of pedestrian crowds in normal and evacuation situations. In M. Schreckenberg, & S. D. Scharma, *Pedestrian and Evacuation Dynamics* (pp. 21-58). Berlin: Springer.

Helbing, D., Molnár, P., Farkas , I. J., & Bolay, K. (2001). Self-organizing pedestrian movement. *Environment and Planning B: Planning and Design*, 361-383.

Hensher, D. A., & Greene, W. H. (2003). The mixed logit model: The state of the practice. *Transportation*, 133-176.

Hermansen, K. (2013). *SMART station: Passenger flow/tracking project.* Vodskov, Denmark: BLIP Systems.

Hermansen, K. (2014, February 14). Data combining and filtering algorithm of SMART station scans. (D. Ton, Interviewer)

Hill, M. R. (1982). *Spatial structure and decision-making aspects of pedestrian route selection trhough an urban environment.* Lincoln, Nebraska: University of Nebraska.

Hoogendoorn, S. P., & Bovy, P. H. (2002). Normative pedestrian behaviour theory and modelling. *Transportation Research Board*, 219-245.

Hoogendoorn, S. P., & Bovy, P. H. (2004). Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B*, 169-190.

Hoogendoorn, S. P., & Bovy, P. H. (2005). Pedestrian travel behavior modeling. *Network and Spatial Economics*, 193-216.

Hoogendoorn, S. P., Daamen, W., & Bovy, P. H. (2001). Microscopic pedestrian wayfinding and dynamics modelling. In M. Schreckenberg, & S. Sharma, *Pedestrian and Evacuation Dynamics* (pp. 123-154). Berlin: Springer.

Hoogendoorn-Lanser, S. (2005). *Modelling travel behaviour in multi-modal networks.* Delft: TRAIL Research School.

Lee, J. Y., Lam, W. H., & Wong, S. C. (2001). Pedestrian simulation model for Hong Kong underground stations. *IEEE Intelligent Transportation Systems Conference Proceedings* (pp. 554-558). Oakland: IEEE.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, 229-254.

Manski, C. F., & McFadden, D. (1981). *Structural analysis of discrete data with econometric applications.* Cambridge: The MIT Press.

Montello, D. R. (1991). Spatial orientation and the angularity of urban routes: a field study. *Environment and Behavior*, 23-47.

Nie, W. (2000). Waiting: Integrating social and psychological perspectives in operations management. *Omega*, 611-628.

Nio, I. (2012, September). *Het station als publieke ruimte.* Retrieved July 2, 2013, from Spoorbouwmeester: www.spoorbouwmeester.nl/beleid/station

NS. (2011). *Keten Informatie Systeem.* Utrecht: NS.

NS. (2014). *Producten*. Retrieved December 23, 2013, from NS: http://www.ns.nl/reizigers/producten

NS Stations. (2013). *Koopgedragonderzoek 2013: High level rapportage.* Utrecht: NS stations.

NS Stations. (2013). *Ontwikkelplan Stations (OPS).* Utrecht: NS Stations.

NS Stations, & Altuïtion. (2013). *ZMET Express NS Stations - Stationsbeleving.* Altuïtion.

Oxford Dictionaries. (2013a). *Defintion of route in English*. Retrieved November 1, 2013, from Oxford Dictionaries: http://www.oxforddictionaries.com/definition/english/route

Oxford Dictionaries. (2013b). *Definition of activity in English*. Retrieved November 1, 2013, from Oxford Dictionaries: http://www.oxforddictionaries.com/definition/english/activity?q=activity+

Parkinson, T., & Fisher, I. (1996). *Rail transit capacity.* Washington: National Academy Press.

Qi, Z., Baoming, H., & Dewei, L. (2008). Modeling and simulation of passenger alighting and boarding movement in Beijin metro stations. *Transportation Research Part C*, 635-649.

Rogerson, J. (2014, January 29). *Bluetooth: What does it really do and will it be replaced?* Retrieved February 4, 2014, from Techradar.phones: http://www.techradar.com/news/phone-and-communications/mobile-phones/what-is-bluetooth-how-it-works-and-how-you-can-use-it-1141428

Roos, M., & Arends-Tóth, J. (2013). *De smartphone als waarneemmethode bij mobiliteitsonderzoek.* Den Haag: CBS.

Root, G. S., & Recker, W. (1981). Toward a dynamic model of individual activity pattern formulation. *International Conference on Travel Demand Analysis*, (pp. 2-22). Oxford, England.

Schadschneider, A., Klingsch, W., Klüpfel, H., Kretz, T., Rogsch, C., & Seyfried, A. (2008). Evacuation dynamics: Empirical results, modeling and applications. In B. Meyers, *Encyclopedia of complexity and system science.* Berlin: Springer.

Seneviratne, P. N., & Morrall, J. F. (1985). Analysis of factors affecting the choice of route of pedestrians. *Transportation Planning and Technology*, 147-159.

Stokols, D. (1972). On the distinction between density and crowding: some implications for future research. *Psychological Review*, 275-277.

Taylor, S. (1994). Waiting for services: the relationship between delays and evaluations of service. *Journal of Marketing*, 56-69.

Thurau, J. (2013). *Datenauswertung - Beobachtungen Laufen.* Bern: SBB CFF FFS.

Timmermans, H. (2009). *Pedestrian behavior: Models, data collection and applications.* Bingley: Emerald Group Publishing Limited.

Train, K. (2003). *Discrete choice methods with simulation.* Cambridge: Cambridge University Press.

Treinreiziger.nl. (2013). *Spoorkaart 2013.* Retrieved December 04, 2013, from Treinreiziger.nl: http://www.treinreiziger.nl/userfiles/image/artikel/pdf/spoorkaart_2013_A4.pdf

van den Heuvel, J., Thiellier, E., & van Gerwen, N. (2013, March). Privacy by design; Bij reizigersmetingen op stations. *Privacy & Compliance*, pp. 17-21.

van Hagen, M. (2011). *Waiting experience at train stations.* Delft: Eburon Academic Publishers.

van Venrooij, T. (2011, Janury 4). *Loopsnelheid voorspelt mortaliteit*. Retrieved February 26, 2014, from Medisch Contact: http://medischcontact.artsennet.nl/archief-6/tijdschriftartikel/89684/loopsnelheid-voorspelt-mortaliteit.htm

Veenendaal, A. J. (2004). *Spoorwegen in Nederland; Van 1834 tot nu.* Boom.

Verlander, N. Q., & Heydecker, B. G. (1997). Pedestrian route choice: an empirical study. *PTRC European Transport Forum* (pp. 39-49). Brunel University, England: Transportation Planning Methods.

Versichele, M., Neutens, T., Delafontaine, M., & Van de Weghe, N. (2012). The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Applied Geography*, 208-220.

Voskamp, A. (2012). *Measuring the influence of congested bottlenecks on route choice behavior of pedestrians at Utrecht Centraal.* Delft: Delft University of Technology.

Vuchic, V. R. (2007). *Urban transit: Systems and technology.* Wiley online Library.

Zhu, W. (2008). *Bounded rationality and spatio-temporal pedestrian shopping behavior.* Eindhoven: Eindhoven University of Technology.


cosmetics